THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

lume XXVIII October, 1937 Number 7

THE VALIDITY OF THE PORTEUS MAZE TEST

MOSHE BRILL

The Hebrew University, Jerusalem, Palestine

Investigators who were previously engaged in the study of intelli-nce, have turned their attention in the past several years to the udy of personality and adjustment. It has been recognized by these vestigators that no one concept, however important, suffices to inter-et all clinical cases. Even such fundamental concepts as general telligence and mental level are by no means sufficient for the diagno-s of all types of maladjustment. There are many individuals who nction on an average mental level, or who, in a manner of speaking, e even bright, but still fail to meet the ordinary requirements placed pon them by the social group. A diagnostic classification, namely, ental age, mental level, or intelligence quotient, does not in itself terpret or solve a person's failure to adjust. Lowrey[1] reports that ild guidance clinics have found mental deficiency to be a lesser factor the production of difficult behavior than was anticipated by ychiatrists.

Although social behavior cannot be sharply divided into two distinct pes, namely, good and bad, individuals can be ranked from best to rst insofar as their social adjustment is concerned. The classroom acher can thus rank his pupils and point out those who do not, for me reason, get along socially. He may also observe that certain pils, though maladjusted at the moment, respond to training and come better adjusted with maturity. There are also those children o do not respond to training and remain quite hopeless insofar as ial adjustment is concerned. Some of these, and other children, are ught to the guidance clinic. Upon studying them, the clinician ds himself in need of some working concepts other than general elligence and mental level by means of which he could better inter-t the dynamic behavior of the maladjusted individual who functions

on a normal, or above normal, mental level. Psychopathy is one of the concepts introduced to satisfy this need.

A survey of the psychological and psychiatric literature on the maladjusted reveals many definitions of the term psychopathy and its synonyms. A careful examination of these definitions indicates that the plausible ones, at least from the clinical viewpoint, are those which describe the behavior of the psychopath and at the same time suggest that the abnormality is due to some cause or causes other than mental level. The term unstable is common in current professional literature. This is probably due to the fact that of all other terms used the word unstable actually describes the overt behavior of the individual usually classified as a psychopath. His behavior is unstable and, therefore, unpredictable. Unevenness of functioning, or sudden change in modes of reacting to similar situations, characterize certain psychopaths best. The significant characteristics and, hence, diagnostic difference between most psychopaths and non-psychopaths is the failure of the former to make a satisfactory social adjustment. He is reported by clinicians to lack, among others, the following: the habit of forethought; the ability to foresee and to weigh the possible consequences for self and others resulting from certain acts; planfulness; prudence; ability to profit by experience; capacity for inhibition; power of adaptation to novel situations.

THE PORTEUS MAZE TEST

Berry and Porteus[2] state: "We have already indicated our approval of Pearson's dictum, that it is the detection of the potential social inefficient which is the chief concern of mental diagnosis. Society is no less injured by the mentally unstable, than by the intellectually feeble-minded." The early diagnosis of the incipient socially maladjusted individuals is thus a primary concern, if not the most important one, of the clinician. To make this diagnosis objective, objective techniques are necessary. It is evident, however, that intellectual retardation is not the salient feature of the psychology of a great many of the socially maladjusted. Tests such as the Binet have, therefore, only a limited value in the study of these individuals. According to Berry and Porteus, "What is evidently required to supplement the Binet examination, are tests of other capacities important for social sufficiency. These social capacities are in the main dependent on the ability to control impulsive action. In a word, we wish to test self-

control." The Porteus Maze test was devised to measure these capacities.

According to Porteus, his Maze test not only measures intelligence, but also other abilities and personality traits that he found to correlate highly with social adjustment. His studies on the reliability and validity of the Maze test are reported in his *Studies in Mental Deviations;*[4] *The Maze Test and Mental Differences;*[5] and others. The following summarizes Porteus' claims for his Maze test: "The Porteus tests represent an attempt to evaluate socially valuable characteristics not fully tested by the Binet. These capacities are mainly prudence, forethought, planning capacity, ability to improve with practice, and adaptability to a new situation. The deficiencies in these respects, even more than in intellectual attainments, distinguish high-grade defectives from normal children; hence the value of the tests for diagnostic purposes. They fulfill the requirements of supplementary tests because they can be standardized and arranged in the form of a scale. They can be easily applied, and they test highly correlated mental functions in simple situations, so that results may be readily interpreted."[6]

A critical study of the original Porteus investigations reveals the fact that his earlier conclusions as to the validity of the test from the standpoint of social adaptation were not fully justified. Porteus' earlier evidences were not conclusive by any means. Garrett and Schneck[7] state: "From the fact that the Mazes correlate somewhat higher than Binet with industrial capacity and social capacity, and from an examination of many individual records, Porteus concludes that his Mazes measure aspects of character and temperament not tested by the Binet. While this may be true, the evidence for it is exceedingly meager. To be sure, Burt considers the Porteus Mazes to be useful supplements in estimating social as distinguished from educational efficiency. But just how he knows that social efficiency is the trait measured by the Maze is not made clear."

In a study by Cornell and Lowden,[8] they concluded in part: "While the groups show some differences of considerable interest, the differences in individual performance are scarcely clearcut or consistent enough to be of more than the slightest value in diagnosis." "The qualitative side of the test is more significant than the quantitative in showing temperamental factors." "A qualitative report of the Stanford examinations shows more factors significant in diagnosis than the Porteus."

THE PROBLEM*

The present study consists of a comparison of the performance of fifty well-adjusted and fifty seriously maladjusted mentally deficient boys on the Porteus Maze test. The purpose of the study is to determine the validity of this test. The three specific questions to be answered in this study are: (1) Does the Porteus Maze test age differentiate reliably between the socially adjusted and the socially maladjusted mentally deficient boy? (2) does the difference between the Binet and Porteus test ages differentiate reliably between the adjusted and the maladjusted mentally deficient? and (3) does scatter, or the amount of variability, on the Maze test differentiate reliably between the adjusted and maladjusted mentally deficient?

The present study is one of a series of studies on the validity of various commonly used tests and scales for differentiating the socially maladjusted mentally deficient.[9]

METHOD AND PROCEDURE

Equating the Two Groups.—The two groups compared were well equated in the following items:

(1) Nationality and race. Each group consisted of forty-nine native white and one colored.

(2) All subjects in both groups were male.

(3) Chronological age.† The means of the adjusted and maladjusted groups were 235.38 ± 4.21 and 235.68 ± 5.06 respectively; the ranges one hundred fifty-nine to three hundred fifty-two and one hundred fifty-seven to three hundred ninety-five respectively; and the sigmas 44.18 and 53.05 respectively. The difference of the means was .30 ± 6.585 and the critical ratio .046.

(4) Binet test age.† The mean test ages of the adjusted and maladjusted groups were 107.94 ± 1.93 and 107.50 ± 2.19 respectively; the ranges seventy-eight to one hundred eighty-two and seventy-two to one hundred eighty-seven respectively; and the sigmas 20.22 and 22.92 respectively. The difference of the means was .44 ± 2.915 and the critical ratio .151.

(5) Number of months in the Colony prior to the investigation. The means of the adjusted and maladjusted groups were 44.00 ± 3.17 and 43.96 ± 3.48 respectively; the ranges nine to one hundred eighty-two and eight to one hundred ninety respectively; and the sigmas 33.26 and 36.48 respectively. The difference of the means was 0.4 ± 4.709 and the critical ratio .009.

---

* This study was made at the State Colony, New Lisbon, N. J.

† In months.

(6) General physical health as rated by the resident physician. Of the boys in each group, fifty-two per cent were rated average, six per cent below average,and forty-two per cent above average.

Comparing the Two Groups in Adjustment.—The outstanding difference between the two groups was in the matter of social adjustment. To study this difference the Vineland Adjustment Score Card[10] was used. In a study by Oosthuizen,[11] he concluded that "the claim that the method is workable and efficient in objectively measuring conduct is fully justified." The mean Adjustment Score Card scores of the adjusted and maladjusted groups were 107.04 ± 1.60 and 59.30 ± 1.99 respectively; the ranges seventy-three to one hundred thirty-eight and twenty-seven to ninety-six respectively; and the sigmas 16.75 and 20.85 respectively. The difference of the means was 47.74 ± 2.553 and the critical ratio 18.70. This critical ratio is more than necessary to prove conclusively the reliable difference between the two groups in adjustment as measured by the Vineland Score Card.

To check further on the reliability of the difference between the two groups in continuous social adjustment, the Elm (detention) cottage records of the boys for the past two and a half years were studied. The institution where the study was made, the New Jersey State Colony for Feeble-Minded Males, New Lisbon, N. J., maintains the Elm cottage for the boys who fail to adjust in the ordinary institutional environment. This cottage is different from the open cottage only in that the boys in it are kept under constant supervision. The

	Adjusted	Maladjusted
1. Total number of days in colony for past two one-half years	40,177	38,942
2. Total number of days in Elm for past two one-half years	471	11,502
3. E.D. index (ratio of two over one)	.012	.295
4. Total number of commitments to Elm	49	299
5. Mean number of commitments per individual	.98	5.98
6. Per cent of boys never committed to Elm	56	0
7. Total number of recommitments to Elm	36	249
8. Mean number of recommitments to Elm	.52	4.98
9. Per cent returned from jails and State Hospitals	0	14.03
10. Per cent committed to Elm for minor offenses	10.20	3.01
11. Ratio between adjusted and maladjusted in three	1	24.58
12. Ratio between adjusted and maladjusted in five	1	6.10
13. Ratio between adjusted and maladjusted in eight	1	9.58

facts in the table concerning the adjustment of the boys in the two groups were brought out from a study of the Elm cottage data (E.D.).

To summarize, an analysis of the Elm, or detention cottage, data has shown clearly the significant difference between the two groups in continuous social adjustment for the past two and a half years.

The Test.—The Porteus Maze Test, Vineland 1933 Revision,[12] was used in the present study. The administering and scoring of the test was in strict accordance with Porteus' instructions. Of the tests for adults, only Adult I was used.

RESULTS

A study of Table I reveals the difference between the two distributions of the Porteus test ages, as well as the overlapping between them. It is sufficient to indicate, for instance, that forty per cent of the adjusted boys and only twenty-six per cent of the maladjusted boys scored below eight years.

TABLE I.—DISTRIBUTION OF PORTEUS TEST AGES OF THE BOYS IN THE TWO GROUPS

Group	Test ages												Total
	4	5	6	7	8	9	10	11	12	13	14	15	
Adjusted	0	4	5	11	6	2	2	5	2	3	6	4	50
Maladjusted	2	3	3	5	2	5	4	9	5	6	2	4	50
Total	2	7	8	16	8	7	6	14	7	9	8	8	100

The Difference between the Porteus Test Ages of the Two Groups.—Forty per cent of the adjusted boys passed the Adult I test on one of the four trials allowed, as did thirty-eight per cent of the maladjusted boys. The difference between the groups in the Adult I test is negligible.

A study of Table II indicates that the mean of the maladjusted is larger than that of the adjusted by 6.36 ± 5.07. The critical ratio being 1.25, it follows that the chances are approximately eighty in one hundred that the maladjusted will always score higher on the Porteus test than the adjusted. However, there is a far greater tendency for the two groups to overlap than to differ. Thirty-four per cent of the adjusted boys scored above the median for the maladjusted group (one hundred thirty-two), and thirty per cent of the maladjusted boys scored below the median for the adjusted group

(one hundred two). There is a marked difference between the mean and medians of both groups. Holzinger's measure of skewness formula, $Sk = \frac{3(M - Md)}{SD}$[13] was applied. For the distribution of the adjusted group the skewness is +1.11; the skewness for the distribution of the maladjusted is −.77. This tendency to skewness explains in part the differences between the means and medians of the two groups.

TABLE II.—PORTEUS MAZE TEST AGES OF THE TWO GROUPS

	Adjusted	Maladjusted
Range	60–180*	54–186
Mean	116.16 ± 3.65†	122.52 ± 3.52
SD	38.30	36.86
Difference of the means		6.36 ± 5.07
Critical ratio		1.25

* In months.
† PE.

To summarize, the difference between the central tendencies of the two groups in Porteus test ages was found to be statistically unreliable. The Porteus Maze test age does not differentiate reliably between the adjusted and the maladjusted mentally deficient boy.

The Difference between the Binet and Porteus Test Ages.—As it is the general practice of psychological clinics to use the Binet scale together with other psychometrics, it was deemed worthwhile to study here the tendencies of the two groups to deviate on the Porteus test from their respective Binet test levels.

TABLE III.—MEAN BINET AND PORTEUS TEST AGES OF THE TWO GROUPS

	Adjusted	Maladjusted
Binet scale	107.94 ± 1.93	107.50 ± 2.19
Porteus test	116.16 ± 3.65	122.52 ± 3.52
$M_{Port.} - M_{Binet}$	9.22 ± 4.13	15.02 ± 4.15
Critical ratio	2.23	3.55

It appears from Table III that both groups scored higher on the Porteus test than on the Binet Scale. There are approximately ninety-three chances in one hundred that the adjusted group will always score higher on the Porteus than on the Binet, while the chances

are ninety-nine in one hundred that the same will hold true for the maladjusted group. This difference between the two groups is, of course, of no practical value to the clinician. Just how much of a difference in years and months, or in terms of any other units, must there be between an individual's Binet and Porteus ratings to cause him to be classified as either adjusted or maladjusted?

An analysis of the data in Table IV indicates that although there appear to be some differences between the two groups in their tendencies to deviate on the Porteus test from their Binet test levels, these differences are small, and are of little or no consequence. Seventy per cent of the maladjusted and only fifty-two per cent of the adjusted scored higher on the Porteus test than on the Binet scale. In the case of only one maladjusted subject were the Binet and Porteus test ages alike. The term deviation stands here for the difference between the Porteus and Binet test ages. Plus-deviation stands for the difference between the two test ages where the Porteus TA is higher than that of the Binet. Minus-deviation stands for the difference between the two test ages where the Porteus TA is lower than that of the Binet. Comparing the two groups in their deviations, it is found that the median plus-deviations of the adjusted and the maladjusted groups are two years seven months and two years four months respectively. The median minus-deviations of the adjusted and maladjusted groups are one year four months and one year eight months respectively. In other words, the maladjusted scored higher on the Binet than on the Porteus more often than did the adjusted. Yet, the difference between the two groups in the amount of deviation between the two tests is insignificant, and, hence, of no value clinically. For instance, an individual who scores two and a half years higher on the Porteus than on the Binet, could not be classified, because of this difference between his two test ages, as either adjusted or maladjusted, as stable or unstable. There is no evidence in this study to support Porteus' contention, namely, that "failure in the tests is more significant than success."[14] To summarize, the difference between the Porteus and Binet test ages is an index of no value to the clinician in the diagnosis of maladjustment.

Correlations.—Porteus reported the correlation of the Porteus Maze test with the Binet to be positive at each age level, ranging for boys, from $.24 \pm .08$ to $.61 \pm .066$, and for girls from $.41 \pm .064$ to $.75 \pm .034$.[15] The average r obtained approximates .51; .47 for boys and .56 for girls. The Pearson product moment coefficient of correlation between the Binet and the Porteus test ages of all one hundred

TABLE IV.—DISTRIBUTION OF THE DEVIATIONS BETWEEN THE BINET AND PORTEUS TEST AGES IN YEARS AND MONTHS OF THE ADJUSTED AND MALADJUSTED

Porteus TA minus Binet TA	Adjusted	Maladjusted	Adjusted plus maladjusted
+7:6	0	1	1
+7:0	0	0	0
+6:6	0	0	0
+6:0	1	1	2
+5:6	3	3	6
+5:0	2	3	5
+4:6	0	1	1
+4:0	3	1	4
+3:6	1	3	4
+3:0	3	2	5
+2:6	3	4	7
+2:0	3	2	5
+1:6	3	5	8
+1:0	2	5	7
+ :6	2	4	6
0	0	1	1
− :6	5	2	7
−1:0	2	2	4
−1:6	7	2	9
−2:0	3	3	6
−2:6	4	1	5
−3:0	1	0	1
−3:6	1	2	3
−4:0	0	0	0
−4:6	1	1	2
−5:0	0	1	1
N	50	50	100
Per cent Porteus TA higher than Binet TA	52	70	61
Per cent Porteus TA lower than Binet TA	48	28	38
Range of deviation between Porteus and Binet TA	−4:6 to +6:0	−5:0 to +7:6	−5:0 to +7:6
Median Porteus TA higher than Binet TA	2:7	2:4	2:4
Q Porteus TA higher than Binet TA	1:3	1:9	1:4
Median Porteus TA lower than Binet TA	1:4	1:8	1:4
Q Porteus TA lower than Binet TA	:10	1:2	:10

subjects in the present study was found to be .52 ± .049; the coefficients of correlation for the adjusted and maladjusted groups separately, were found to be .60 ± .061 and .46 ± .075 respectively. Here again, it is apparent that the difference between the adjusted and the maladjusted groups in their performance on the Porteus Maze test is not of sufficient magnitude to be of any practical value in clinical diagnosis.

Scatter in the Porteus Test.—With regard to scatter in the Maze test, Porteus states: "As a general conclusion, it may be stated that scattering in the Porteus test is characteristic of certain psychopathic and delinquent children somewhat over-rated by the test. . . . Scattering is not characteristic of all the psychopathic or delinquent as it frequently happens that children of these types have already scored very low in the Maze test in comparison with other tests."[16]

TABLE V.—NUMBER OF YEARS OF SCATTER IN THE PORTEUS MAZE TEST OF THE TWO GROUPS

Groups	Number of years										Total
	1	2	3	4	5	6	7	8	9	10	
Adjusted	0	4	9	6	3	5	1	0	7	15	50
Maladjusted	3	2	6	2	3	2	2	0	9	21	50
Total	3	6	15	8	6	7	3	0	16	36	100

In Table V the distribution of number of years of scatter in the Porteus test by the two groups is given. From this table it appears that the difference between the two groups in the amount of scatter in the Porteus test is statistically unreliable. The amount of overlapping of the two distributions is significant. Thirty per cent of the Adjusted boys had a scatter on the Porteus test above the median for

TABLE VI.—NUMBER OF YEARS OF SCATTER IN THE PORTEUS MAZE TEST BY THE TWO GROUPS

	Adjusted	Maladjusted
Range	2–10	1–10
Mean	6.48 ± .29	7.30 ± .30
SD	3.05	3.18
Difference of the means	.82 ± 1.32	
Critical ratio	.62	

the Maladjusted group (nine), and thirty-two per cent of the Maladjusted boys had a scatter above the median for the Adjusted group (six).

It appears from Table VI that the differences between the mean number of years of scatter in the Porteus test of the two groups is statistically unreliable. The critical ratio being .62, it follows that the chances are sixty-six in one hundred that the Maladjusted will always have a large scatter in the Porteus test than the adjusted. Scatter in the Porteus does not reliably differentiate the adjusted from the maladjusted, and is, hence, no index of diagnostic value as regards adjustment.

SUMMARY AND CONCLUSIONS

A comparative study was made of the performance of fifty socially well-adjusted and fifty seriously maladjusted mentally deficient boys on the Porteus Maze test. The purpose of the study was to determine the validity of this test. The three specific points to be covered in this study were: (1) Does the Porteus Maze test differentiate the adjusted from the maladjusted? (2) is the difference between the Porteus and Binet test scores of diagnostic value as regards adjustment? and (3) does the scatter on the Maze test differentiate reliably between the adjusted and maladjusted mentally deficient boy?

The two experimental groups, both living in a state institution, were well equated in race and nationality, sex, chronological age, Binet test age, number of months in the institution prior to the study, and in general physical health as rated by the resident physician. The only outstanding difference between the two groups was in the matter of continuous social adjustment for at least the past two and a half years. The adjustment of the two groups was studied by means of the Vineland Adjustment Score Card, and also by an analysis of the Elm detention cottage records of the subjects in the two groups. The following conclusions may be drawn from this study:

1. The maladjusted group scored higher on the Porteus Maze test than the adjusted group. Although the difference of the means of the two groups is statistically unreliable, it is safe, nevertheless, to question the earlier claims advanced by Porteus that the Maze test does differentiate the socially maladjusted, and that the socially maladjusted score lower than the adjusted on this test. The Porteus test age arrived at by the method of administering and scoring described in the 1933 Revision, does not reliably differentiate the adjusted from the mal-

adjusted. The adjusted actually scored higher on the Maze test than did the maladjusted.

2. There are more chances for the maladjusted than for the adjusted to score higher on the Porteus test than on the Binet scale. Yet, the difference between the Binet and Porteus test ages was found not to differentiate between the adjusted and the maladjusted groups. The difference between the two tests is, therefore, no index of diagnostic value as regards adjustment. There is no evidence in this study to support Porteus' contention, namely, that "failure in the tests is more significant than success."

3. The amount of scatter on the Porteus Maze test does not differentiate reliably between the adjusted and the maladjusted, and is, hence, of no clinical value in the diagnosis of maladjustment.

4. Both the Adjusted and the Maladjusted scored higher on the Porteus test than on the Binet scale. The tendency for the Maladjusted to score higher on the Porteus test than on the Binet scale is stronger than that for the Adjusted. The coefficient of correlation between the Binet and Porteus test ages of the Adjusted was found to be higher than that of the Maladjusted. These two findings indicate that the relationship between the Binet and Porteus test ages is more definite in the case of the Adjusted than in the case of the Maladjusted. However, this difference between the two groups is too vague to be of any practical value.

The preceding conclusions should not be misinterpreted to the effect that the test has no clinical values whatsoever. On the contrary, the clinician might use this test very profitably as a means for a qualitative study of the individual's reactions under standardized procedure. It is desirable that additional investigations of these clinical values should be carried out.

REFERENCES

1. The Committee on Physically and Mentally Handicapped, the White House Conference on Child Health and Protection, *The Handicapped Child.* New York: The Century Company, 1933, p. 389.
2. Berry, R. A. and Porteus, S. D.: *Intelligence and Social Valuation.* Vineland, N. J.: The Training School, Department of Research, No. 20, May, 1920, p. 83.
3. *Ibid.*, p. 67.
4. Vineland, N. J.: The Training School, Department of Research, No. 24, Oct., 1922.
5. Vineland, N. J.: Smith Printing and Publishing House, 1933.
6. Berry and Porteus: *Op. cit.*, p. 91.

7. *Psychological Tests, Methods, and Results.* New York: Harper and Brothers, 1933, Part II, pp. 84–85.
8. Cornell, E. L. and Lowden, G. L.: "A Comparison of the Stanford and Porteus Tests in Several Types of Social Inadequacy." *Jour. Abnor. Psych.*, Vol. XVIII, 1933, pp. 41–42.
9. Brill, M.: *A Comparative Study of the Performance of Adjusted and Maladjusted Mentally Deficient Boys on Twenty-Two Tests and Scales.* Ph. D. Thesis, New York University, 1935.
———: "Psychometric Data as Indices of Instability." *Proc. and Addr. Amer. Asso. Ment. Defic.*, Vol. XL, 1935, pp. 421–434.
———: "Performance Tests as Aids in the Diagnosis of Maladjustment." *Jour. Genet. Psych.*, Vol. XLIX, 1936, pp. 199–214.
10. Yepsen, L. N.: "The Measurement of Social Behavior." *Proc. and Addr. Amer. Asso. for the Study of the Feeble-minded*, Vol. XXXIII, 1928, p. 124.
11. "Efficiency of the Vineland Adjustment Score Card for Measuring Social Behavior." *Jour. Educ. Res.*, Vol. XXIII, April, 1931, pp. 280–287.
12. Porteus, S. D.: *The Maze Test and Mental Differences. Op. cit.*
13. *Statistical Methods for Students of Education.* Boston: Ginn and Company, 1928, p. 123.
14. Berry and Porteus: *Op. cit.*, p. 69.
15. *Ibid.*, p. 74.
16. Porteus, S. D.: *Studies in Mental Deviations. Op. cit.*, p. 112.

REPEATED TESTS WITH THE CAVD SCALE

RUDOLF PINTNER AND MILDRED STANTON

Teachers College, Columbia University

The curve for the general growth of intelligence from birth to maturity is usually assumed to be a parabolic curve. Intelligence increases rapidly during infancy and early childhood and then gradually slows down during adolescence and ceases during early maturity. Some psychologists, however, have argued for a straight line curve during the elementary-school period, basing their arguments upon the equal annual increments of gain in score on group tests of intelligence. But scores on almost all group intelligence tests have not been equated for difficulty, so that it is impossible to interpret the significance of gains in score from one year to the next. The Thorndike CAVD test is the only intelligence test so far constructed which has attempted to make each unit on the scale equal in difficulty to every other unit. A gain of one unit on this test may, therefore, be considered as a definite gain in altitude of intelligence, regardless of the level of difficulty where this gain takes place. The results of some annual tests with this test will be presented.

THE POPULATION TESTED

The CAVD Test was given annually to a number of children. Table I shows the grade location of these cases at the time of the last test and the number of annual tests given. Practically all the

TABLE I.—NUMBER OF CASES TESTED

Number of times tested	Grade at the time of last test												Total
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	
2	3	0	0	9	3	5	2	9	..	1	..	..	32
3	..	3	1	1	13	10	2	7	..	..	..	2	39
4	..	..	1	1	0	14	7	6	..	..	..	3	32
5	..	..	..	..	..	..	6	21	..	..	..	..	27
6	..	..	..	..	..	..	..	10	..	..	..	..	10
Total.....	3	3	2	11	16	29	17	53	..	1	..	5	140

tests were given eleven, twelve or thirteen months apart. A total of one hundred forty children were tested with varying numbers

carried from two to six years. Most of the testing took place while the children were in grades I to VIII. The children attended public schools in New York City in good residential neighborhoods and would on the average test slightly above IQ one hundred.

ANNUAL GAINS OR LOSSES ON ALL RE-TESTS

We have a total of three hundred sixty-four gains or losses from one year to the next, taking into account every difference from one test to the next annual test. These three hundred sixty-four gains or losses are distributed as follows:

Gains or Losses		*N*
Plus	3.5 to 3.9	1
	3.0 to 3.4	4
	2.5 to 2.9	9
	2.0 to 2.4	19
	1.5 to 1.9	46
	1.0 to 1.4	101
	.5 to .9	118
	0 to .4	54
Minus	.1 to .5	8
	.6 to 1.0	3
	1.1 to 1.5	0
	1.6 to 2.0	1
Total		364
Average gain		1.06
SD		.71
Extreme range		−1.8 to +3.5
Middle 50 per cent		+0.5 to +1.4
Middle 80 per cent		+0.2 to +1.9

The average annual gain is 1.06 points. Only twelve or 3.4 per cent show losses. One of these losses is very large, namely 1.8 points. Fourteen of the gains are very large, being 2.5 points or more. The great majority of the gains, however, range from one-half to one and one-half points. This total distribution of gains contains children who have been tested only twice as well as children who have been tested up to six times. Repeated annual testing may produce practice effects and lead to larger gains. It would be well, therefore, to study the gains according to the number of tests taken.

Table II show these mean annual gains. The first row shows that ten children were tested each year over a period of six years and that the average gain from the first to the second year (test 2 minus test 1) was 1.26 points; the next annual gain (test 3 minus test 2) was

TABLE II.—AVERAGE ANNUAL GAINS PER NUMBER OF TIMES TESTED

Number of tests	*N*	Tests taken				
		2 – 1	3 – 2	4 – 3	5 – 4	6 – 5
6	10	1.26	1.03	0.48	0.97	1.06
5	27	1.36	1.07	0.79	0.75	
4	32	0.91	1.02	0.92		
3	39	0.99	1.00			
2	32	1.24				

1.03 points; and so on. A study of all these annual gains shows no indication of increasing gains from year to year such as might be expected if the tests were very susceptible to the effects of practice. The CAVD test is evidently similar to other intelligence tests in this respect. The interval of a year between tests is sufficiently long so that no obvious practice effects remain. The results in Table II also indicate that there is no constant tendency for the gains to increase or decrease from year to year. With remarkable consistency, the gains fluctuate around one point. This table tells us nothing about the annual gain from one chronological age to the next, because the children included in each group vary in chronological age.

GAIN AND CHRONOLOGICAL AGE

If now we take all our gains and losses and distribute them according to the chronological age of the child when such gains were made, we arrive at the average gains shown in Table III. The mean gain

TABLE III.—SHOWING AVERAGE GAINS FOR EACH YEAR'S GROWTH IN CA

CA	7 to 8	8 to 9	9 to 10	10 to 11	11 to 12	12 to 13	13 to 14
Mean gain.....	1.23	1.28	1.00	0.87	0.82	0.84	0.60
Median gain.....	0.9	1.1	1.1	0.8	0.9	0.8	0.7
N.....	11	74	84	74	56	30	10

seems to indicate a rather consistent tendency to decrease as the child grows older. The median gain also indicates the same thing. This tendency is, however, not very marked and the decrease, if any, is very slight. Furthermore, these gains are not based upon the same children at each age. As the numbers are not too great at all

ages, the influence on the means of any particularly bright or stupid children in any one category may conceivably distort the general trend of the means.

If, therefore, we study the gains for each chronological age for such groups as took the same number of tests, we shall keep our group constant from year to year. This cuts down the number of usable cases materially. These mean gains are shown in Table IV. The

TABLE IV.—MEAN GAINS PER CA OF CASES HAVING SAME NUMBER OF RE-TESTS

Number of tests	Number of cases	CA			
		8 to 9	9 to 10	10 to 11	11 to 12
5	26	1.45	0.97	0.84	0.86
4	43	1.38	1.07	0.84	
4	13	....	0.74	1.02	0.75
3	50	1.32	1.09		
3	18	....	0.89	0.99	

first row of this table shows that twenty-six children were tested five times, and all of them were given the first test when eight years old. The second row shows that forty-three children were given four tests, beginning with the first test at year eight. And so on for the other rows of the table. Three rows in this table, having the largest sampling of cases, show a consistent tendency to a slight decrease in score with chronological age. The other two rows show no such tendency, but the number of cases in both instances is small. In general the table seems to point to a probable slight annual decrease in gain at least from ages eight to eleven.

GAINS OF BRIGHT AND DULL CHILDREN

We may raise the question as to whether those scoring high or low on the first test make similar gains. For this purpose we shall take the largest group of cases tested annually beginning at the same chronological age. This is the group of fifty-nine children tested first at age eight and then given two subsequent annual tests. We can divide this group into three groups, high, medium, low, of twenty, twenty and nineteen according to their first score. These three groups and their two average annual gains are as follows:

Group	Scores on first test	Average gains	
High	34.2 to 36.0	0.77	1.35
Medium	32.8 to 34.1	0.89	1.09
Low	30.4 to 32.7	2.23	0.82

This shows that the high group, having scores on the first test from 32.2 to 36.0, made a gain of 0.77 during the first year, and a gain of 1.35 during the second year. Similarly for the other two groups. If this small sample is at all representative, it would seem to indicate that dull children lose very definitely from year to year, whereas bright children tend to increase their gains slightly.

The correlations between tests for these fifty-nine cases are as follows:

$$\text{Test 1 with Test 2, } r = .680 \pm .047$$
$$\text{Test 1 with Test 3, } r = .756 \pm .038$$
$$\text{Test 2 with Test 3, } r = .824 \pm .028$$

Although the correlations are fairly high, it can be seen that there must be many individual fluctuations from year to year in test score.

If we take the twenty-six cases of eight-year-olds tested five times and divide them into two groups of thirteen according to their scores on the first test and then calculate the average annual gains, we have the following results:

Group	Scores	Average annual gains			
High	32.8 to 35.4	0.65	1.08	0.96	0.87
Low	30.4 to 32.7	2.18	0.87	0.72	0.92

Here the brighter cases increase their gain during the second year, but this increase does not continue during the next two years. The gains for the duller cases fall sharply during the second year, and somewhat less so thereafter. The difference between the dull and the bright is very marked during the first and second years but much less so thereafter.

INDIVIDUAL CASES

The greatest number of annual tests given was six. We have ten children who were given six tests, thus giving five annual gains.

These ten cases are shown in Table V. These ten cases show the extreme individual differences that we encounter in the mental growth of children in general. Only two cases show losses, but there is no regularity in the individual gains from one year to the next. Large gains one year may be followed by small gains during the next year and *vice versa*. No case shows a consistent trend toward decrease or increase in gain from one year to the next. Child *C* shows the largest average annual gain, namely 1.36. Child *J* shows the smallest, namely .72. *C*'s advantage over *J* is largely due to one large gain made during the first year.

SUMMARY

One hundred forty children were tested every year on the CAVD Test. The number of annual tests given each child varied from two to six.

TABLE V.—INDIVIDUAL GAINS DURING FIVE YEARS

Years	Children									
	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
1	1.5	1.6	2.9	1.6	2.9	1.6	0	.4	.8	−0.7
2	3.0	.4	1.0	0	.1	.6	1.6	1.1	1.2	1.3
3	−1.8	.4	.5	1.1	.6	.2	.4	1.2	.5	1.7
4	.7	1.4	.8	.7	.9	1.5	1.1	.8	1.3	.5
5	1.0	1.3	1.6	1.1	.8	.8	1.2	.8	1.2	.8
Average..	.88	1.02	1.36	.90	1.06	.94	.86	.86	1.00	.72

The average of all annual gains is 1.06 points on the CAVD Scale. The great majority of gains ranges from one-half to one and one-half points. This average gain of about one point is about one fortieth of the total range of the CAVD Scale. It is about one twentieth of that part of the scale which stretches from what imbeciles can accomplish to what the upper quarter of college graduates accomplish. It is equivalent, as Thorndike says, to about one and one-half years of mental age in the age range from six to twelve.

There seems to be a slight decrease in the average gain from year to year. This is shown to extend from age eight to fourteen in Table III, and also in Table IV for most cases from age eight to twelve.

The difference in the trend of the average annual gains between those who made initial high scores and those who made low scores has been studied. Those who scored low made relatively large gains during the first year followed by a marked drop in gains thereafter. Those who scored high made their largest gain during the second year. It is doubtful whether this represents a general tendency for those who score high and those who score low in this test.

The correlation between tests for fifty-nine cases given three tests ranges from .68 to .82.

Individual gains from year to year fluctuate widely and few individuals, if any, show a regular decrease in gain from one year to the next.

Our results in general seem to point to a slight average decrease in gain and thus to fit in with the theory of a parabolic rather than a straight line curve as representing the general growth of intelligence.

THE BASIS FOR THE PERSONAL CONSTANT

KATHERINE PRESTON BRADWAY AND E. LOUISE HOFFEDITZ

Department of Research, The Training School at Vineland, N. J.

In 1924 Hugo Heinis first proposed the *personal constant* (PC) as a substitute for the *intelligence quotient* (IQ).[4] The concept of the PC has attracted so much attention in America that a translation of the French article which contains the basis for the development of the PC would be valuable to students of mental measurement.* We have translated the original thirty-one pages and have selected from these the part which seems most pertinent to an understanding of the personal constant.

Before presenting this translation it is desirable to go one step further back. In 1922 G. Vermeylen[13] developed a scale for measuring intelligence based on an adaptation of tests already in use. That scale consisted of fifteen "tests," each of which was composed of ten items of one type graduated in order of increasing difficulty. With this scale Vermeylen examined sixty school children ranging in age from six to twelve years. It was the results of these examinations which Heinis treated statistically for the development of the formula for the personal constant. Consequently, a full appreciation of Heinis' study requires some knowledge of the Vermeylen Scale. We are presenting in summary a translation of Vermeylen's description of his scale before offering a translation of Heinis' article.

VERMEYLEN SCALE

1. *Pointing.*—*S* (subject) is shown, one at a time, ten cards which have been marked off into squares, certain of which contain black dots. On the removal of the card, *S* indicates the position of the dots on similar blank squares. The cards range in difficulty from number one, which contains nine squares, one of which has one dot, to number ten which contains sixteen squares, seven of which have one dot each.

2. *Piercing.*—*S* is shown, one at a time, five wooden boards each of which contains one hundred holes variously marked. He is told to pierce with a stylus certain types of holes. The ten degrees of the series range from piercing all the holes of a board to piercing every third hole of a certain type in a board containing five types of holes.

* A subsequent article was published in English by Heinis, but this article does not include the basic considerations which the French article does. See [5].

3. *Repetition* (*Immediate Memory*).—*S* is shown a series of ten cards on which are drawn objects or animals with corresponding names beneath, the number increasing from one to ten. *S* reads them over three times with *E* (examiner) pointing to them. *S* is then requested to repeat the names from memory.

4. *Memory.*—*S* is presented successively with ten familiar objects until he can name them from memory. Forty-eight hours later he is asked to repeat the series.

5. *Identification.*—*S* is shown, one at a time, ten test cards on which are represented abstract designs, the number increasing from one to ten. After a test card is exposed for one second, *S* is shown one of a series of ten master cards, each containing the design or designs of the correspondingly numbered test card together with other designs. The total number of designs on master cards ranges from four to twenty-four. *S* is to designate on the master card the designs which appeared on the test card.

6. *Inversion* (*Reverse Repetition*).—*S* is shown, one at a time, a series of ten words which increase in length and structural complexity. *S* pronounces and spells the exposed word. The card is removed and *S* is asked to spell the word backwards.

7. *Association* (*Treffer Method*).—This series consists of ten sets of paired concrete designs or abstract signs, one of each pair red and the other black. *S* is shown successively the pairs of each set and is then shown a card on which the designs are reproduced, the reds on one side and the blacks on the other. *S* is supposed to couple them as they appeared in the original pairs. The number of pairs in successive sets increases from two to six.

8. *Story Arrangement.*—*S* is shown, one at a time, ten sets of pictures. He is asked to arrange the pictures of each set in a logical order and relate the story they depict. The number of pictures of the ten sets varies without consistent increase. The increasing difficulty of the stories is based on the abstract quality of the ideas depicted.

9. *Absurdities.*—*S* is asked to point out the absurdities in each of ten pictures. The absurdities increase in complexity.

10. *Problems.*—Ten arithmetical reasoning problems are presented (in print) in order of increasing difficulty, with spaces provided for writing the answers.

11. *Analysis.*—*S* is asked to imagine a specified familiar object and to name a designated number of its principal parts. The progressive

difficulty of this series depends upon the increasing complexity of the objects to be imagined and the number of parts to be enumerated.

12. *Naming.*—*E* names three species for which *S* must name the genus (animal, vegetable, clothing); or, *E* states an action which *S* must qualify.

13. *Reconstruction.*—*S* is required to complete a series of ten incomplete drawings which increase in difficulty according to the complexity of the picture and the amount of completion necessary.

14. *Practical Solutions.*—Ten situations are described which present increasingly difficult practical problems for *S* to solve.

15. *Puzzles.*—This test requires the fitting together of five picture puzzles, the reproduction of three geometric designs, and the arrangement of two Healy and Fernald formboards.

S receives a score on each of the fifteen tests according to the number of the most difficult item passed. Since there are ten items in each test, the total possible score is one hundred fifty points. The graduation in the order of increasing difficulty in a series of items was done first purely *a priori* and on the basis of simple probabilities. It was necessary, of course, to make some changes after giving a number of trial tests in order to maintain the graduation.

TRANSLATION FROM HEINIS

[The following is a translation of the first part of Heinis' article in which he evaluates and applies the procedure and data presented in the article by Vermeylen from which the preceding scale was abstracted.]

What interests us particularly in this study* are the results obtained on the sixty normal children between six and twelve years, because we find realized here experimental conditions which may serve our purpose. The number of the subjects is small, but the quality of the data largely compensates for their meagerness. The division of the examination into two sittings separated by a forty-eight-hour interval, and the use of a total testing period three times as long as that which Binet believed sufficient for this kind of examination, are by themselves signs of good precision. Furthermore, the children were chosen from a homogeneous social environment which favored average mental "growth." The children attended the same school and came from families of the small bourgeoisie in the center of Paris. The author (Vermeylen) has chosen only ten subjects from each grade with no concern for their school

* Heinis refers to Vermeylen's study.—Translators' note.

success, with the sole condition "that they have the mean age of the class."

He speaks further of the scale in these terms:* "Following the plan of the Yerkes-Bridges 'point scale' we have laid out a mean scale of points obtained for each child in the fifteen series of graduated items. Then we have calculated the mean of these points for all the children of the same age."

We cannot subscribe to the use of this last procedure on so small a number of subjects because the means are too much affected by the extreme cases. It would certainly be better to express the central tendency by the median. It is fortunate that Dr. Vermeylen has illustrated his article by the graphs of each of the sixty subjects so that we are able to make certain corrections. To facilitate comparison we have tabulated the means taken from Vermeylen's Graph 1 and below them have added the medians.

As shown here by the "differences," the two results do not agree entirely; it would be useless then to push the precision to the point of entering the decimals as Dr. Vermeylen has done.

Considering the general contour of the individual graphs the author (Vermeylen) concludes: "As the children progress in age, the curves become more and more regular and tend toward the horizontal." This is only apparently so and suggests that the graduation of the tests has been artificially stopped at year ten. The method does not permit the differentiation of the scores of older subjects. To prove this it is sufficient to construct a frequency curve for each age based on the number of points obtained by the students for each of the fifteen tests. It is Graph 9, pages 112–123, of Vermeylen which has been used for the elaboration of Table II and the Figs. 1–6.†

Table II shows by the increase at ten that the conditions are satisfactory for measuring the ages six to ten years, and that they are satisfactory for all but the extreme cases between ten and eleven years, but that they are insufficient beyond that age. Among the seventy scores of ten there are a good number which would have been higher, eleven, twelve, or thirteen, had such scores existed. In view of the score distribution at age eleven to twelve (Table II), this excess can approximately be calculated and the median corrected. It is easy to deduce from the trend of the score distribution for each age that the mode is nearer nine than ten, and it follows that the number of cases n corresponding to the score ten is confined

* Vermeylen: *Op. cit.*, p. 50 (Heinis' footnote).

† These figures are not presented here since the data are directly obtainable from Table II. Further translation has been modified accordingly.—Translators' note.

TABLE I.—(HEINIS)

Age	6–7	7–8	8–9	9–10	10–11	11–12
Mean	58	74	99	112	130	138
Median	56	80	99	113	130	138
Differences	2	6	0	1	0	0

within the limits

$$27 < n < 48.$$

The excess would be about thirty and would be distributed under the scores eleven, twelve, and even thirteen, which would augment by 3.5 points the score obtained by each of the subjects from eleven to twelve years. The true central tendency would be represented by the median 141.5 and not 138. (See Table I.)

TABLE II.*—SCORE IN TESTS (HEINIS)

Ages	0	1	2	3	4	5	6	7	8	9	10	11	12
6–7	5	12	22	23	31	27	24	6					
7–8	..	1	5	13	35	35	29	28	2	2			
8–9	..	..	..	..	4	17	47	51	26	5			
9–10	..	..	..	..	..	5	28	45	36	28	8		
10–11	..	..	..	..	..	3	4	11	42	52	38		
11–12	..	..	..	..	..	..	..	5	27	48	70		
11–12 as extended	..	..	..	..	..	..	..	5	27	48	42	22	6

* With ten subjects at each age level taking the fifteen tests, there are one hundred fifty scores at each level. Their values range from zero to ten and are here tabulated according to frequency.—Translators' note.

This established, one could construct a graph of the six median values. Let us imagine the sum of the points on the ordinate and the ages on the abscissa. For simplification, the numbers indicating the years could be reduced to decimal numbers multiplied by ten; thus six and five-tenths years would be represented by the the number sixty-five. This suffices to fix the points as shown in the table at the top of page 506.

This sequence presents a regular curve. Which one? For simplification, we are omitting all the mathematical deductions which were necessitated in these investigations and are limiting ourselves only to the indication of the result. The curve is represented by the logarithm

$$y = a + be^{\frac{x}{d}}$$

	Age	Age number	Median
P_6	6.5	65	56
P_7	7.5	75	80
P_8	8.5	85	99
P_9	9.5	95	113
P_{10}	10.5	105	130
P_{11}	11.5	115	141.5

where a, b, and d represent constant values and e is the base of the natural logarithm. To determine the value of the three constants, it is advisable to choose those points which are the most exactly defined, that is to say, those where the median and the mean coincide, which are P_8 and P_{11}. Let us add to these the most distant point, P_6, and a very simple calculation results in the values:

$$a = 218.5, \qquad b = -429.0, \qquad d = -66.75$$

Then the three points are situated on the logarithm

$$y = 218.5 - 429e^{-\frac{x}{66.75}}$$

Let us now find how the other points, P_7, P_9, and P_{10}, lie with respect to the curve. Let us calculate the values of y for $x = 75$, $x = 95$, and $x = 105$, and compare these with the values experimentally found.

TABLE III.—NUMBER OF POINTS (HEINIS)

Age	Experimentally found	Calculated	Difference	Per cent
7– 8	80	79	1	1.2
9–10	113	115	2	1.7
10–11	130	129.5	0.5	0.4

In Table III we note that the calculated value is one point lower for P_7 and five-tenths lower for P_9. But these are negligible quantities if one considers the small number of subjects on which the mean has been established. Besides, a simple inspection of Table II will satisfactorily confirm this point of view. In particular the row for children seven to eight years and, even more so, that of children nine to ten years shows less homogeneity than the four others, so that the values seventy-nine and one hundred fifteen agree at least as closely as do the medians

with the central tendency of the two ages. We cannot, then, doubt these theoretically found values, and we can conclude *that the above formula is the most probable expression of the development of intelligence between six and twelve years.*

The consideration of the curve which represents the logarithm outside of these limits leads to discussion. Two parts are distinguished here: The part corresponding to the ages four and five-tenths to twelve years, that which we may call "empirical" since the results can be obtained from the Vermeylen data, and the part corresponding to the other years, or the "theoretical" one.

To judge by the empirical segment, it is only above four and five-tenths years that the child is capable of passing any of the tests. Applied to subjects of five years the method leads to more appreciable results. It ceases to exist for ages greater than eleven years.

The mathematical formula permits us to anticipate the theoretical part by the simple interpretation of the following values:
for

$$x = 45, \qquad y = 0$$

for

$$x = 0, \qquad y = -210.5$$

That is to say, no item can be normally passed by children less than four and five-tenths years of age; the first one is already too difficult for them. Then, to give a method accessible for that age, we must complete each series of tests by items still easier. It is necessary to determine their number. Let us suppose that the scores obtained by the Vermeylen procedure are scores spaced by psychologically equal distances—that which in the main can be admitted. All the theoretical scores are indicated as being spaced by psychologically equal distances. Then the 210.5 points which correspond to the ages from birth to four and five-tenths years, and which are lacking, can be obtained by adding to each of the fifteen tests a series of $\frac{210.5}{15} = 14$ items graduated according to increasing difficulty. The score one of Vermeylen would thus become on the average the score fifteen in the new series of items constituting each test. This same reasoning applied to the number of theoretical points to be attributed to the ages superior to twelve years shows us that the graduation of the same tests can be continued to the height of $\frac{218.5 - 141}{15} = 5$ items marking psychological distances

equal to the preceding. The new test would comprise not ten, but at least 14 + 10 + 5 = 29 items. Psychologically speaking a measure thus elaborated would constitute a standard test of absolute and fixed graduation and applicable to all ages.

After reaching these conclusions, we must find a mathematical criterion for the rational scaling of the tests. Let us return to examine the progress of the spiral development of intelligence by the aid of Fig.

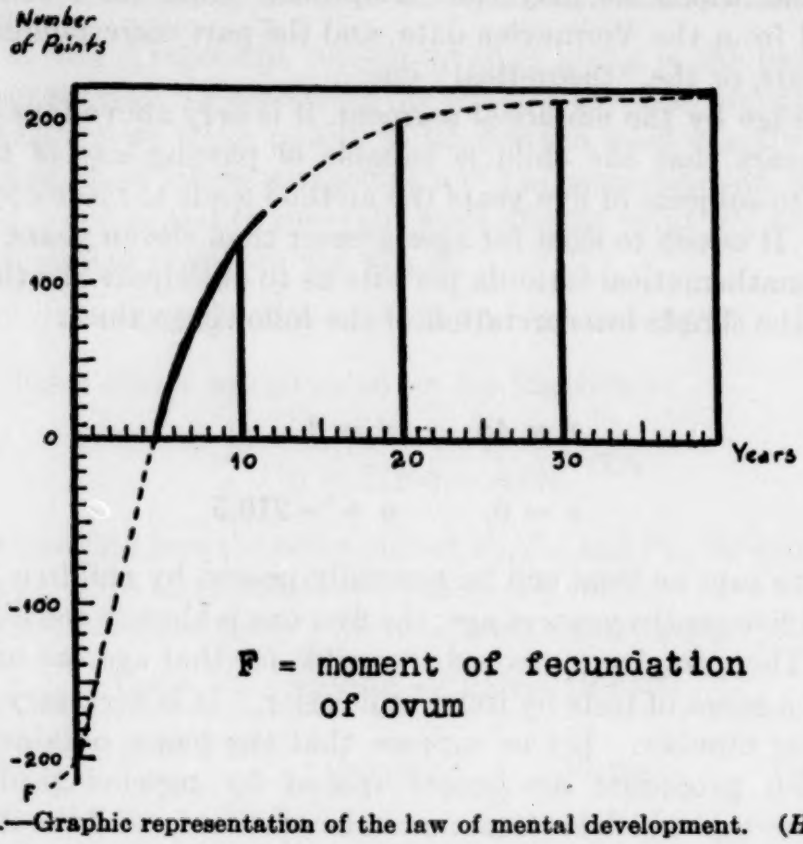

FIG. 7.—Graphic representation of the law of mental development. (*Heinis.*)

7. From the more ample information, we can sum up in Table IV the values of y for the mean of each age and show the differences between consecutive ages. In the fourth column the same y's are expressed in absolute degrees.

To judge from these results, human intelligence does not normally cease to develop during life from birth up to the moment when the destructive action of old age begins. However, the progress, although remaining positive, becomes less and less, until finally at twenty-five years development is practically achieved. Beyond this year and for the rest of life we travel no greater a psychological distance than that achieved in the single year from twelve to thirteen years.

By the graphical method of Fig. 7 we have established both the immense pychological evolution that we achieve during the first five years of our existence (half of the total development)[2] and the constancy of intellectual maturity up to forty years. A segment of especial value is revealed—that which corresponds to the period of puberty—which was until now unknown.

TABLE IV.—(HEINIS)

Age	y	Difference	Absolute degree	Age	y	Difference	Absolute degree
0	−210.5	....	0	20–21	198.6	2.8	409.1
0– 1	−179.5	55.4	30.9	21–22	201.4	2.4	411.9
1– 2	−124.1	47.7	86.3	22–23	203.8	2.0	414.3
2– 3	− 76.4	41.1	131.0	23–24	205.8	1.8	416.3
3– 4	− 35.4	35.3	175.1	24–25	207.6	1.5	418.1
4– 5	− 0.1	30.4	210.4	25–26	209.1	1.3	419.6
5– 6	30.3	26.2	240.8	26–27	210.4	1.2	420.9
6– 7	56.5	22.5	267.0	27–28	211.6	0.9	422.1
7– 8	79.0	19.5	289.5	28–29	212.5	0.8	423.0
8– 9	98.5	16.6	309.0	29–30	213.3	0.8	423.8
9–10	115.1	14.4	325.6	30–31	214.1	0.6	424.6
10–11	129.5	12.3	340.0	31–32	214.7	0.6	425.2
11–12	141.8	10.7	352.3	32–33	215.2	0.4	425.7
12–13	152.6	9.2	363.0	33–34	215.6	0.4	426.1
13–14	161.7	7.9	372.2	34–35	216.0	0.3	426.5
14–15	169.6	6.8	380.1	35–36	216.3	0.3	426.8
15–16	176.4	5.9	386.9	36–37	216.6	0.2	427.1
16–17	182.3	5.0	392.8	37–38	216.8	0.3	427.3
17–18	187.3	4.4	397.8	38–39	217.1	0.2	427.6
18–19	191.7	3.7	402.2	39–40	217.3	0.2	427.8
19–20	195.4	3.2	405.9	40–∞	218.5	...	429.0

In order to be more precise we may say that the *trend of mental development is that of the logarithmic curve*

$$y = a - be^{-\frac{x}{d}}$$

where a, b, d are constant and positive values, and e is the base of the natural logarithm. By the aid of an easy transformation, y can be expressed according to the formula

$$y = a - bQ^{-x}$$

Let i be a particular age and $(i + 1)$ be the following age, then

$$y_{i+1} - y_i = a - bQ^{-(i+1)} - (a - bQ^{-i})$$
$$= bQ^{-i} - bQ^{-(i+1)}$$

It is this last equality which we use to establish the law of the development of intelligence: *The quantitative development of intelligence from one year to the next is equal to the difference between two consecutive terms of a geometrically increasing progression.*

TRANSLATORS' DISCUSSION

In presenting the scale of fifteen tests, ten items each, Vermeylen indicated that the items of each test progressed according to increasing difficulty. The basis of such an increase is implied in his discussion of graduating the items when he says that the *a priori* arrangement gave way to changes after a number of trial tests. Just what methods were used cannot be determined and how strictly his criteria were followed is not known. Inspection of the items of each test throws two points in question:

1. the steepness of the increasing difficulty of items;
2. the equality of steps between items.

We have no way of knowing that these points were not amply considered when the scale was set up, but lack of information along with the very nature of the items leads us to believe that the steepness varies markedly and that the steps are considerably unequal. Change in steepness of item increase from test to test is acceptable if it is accompanied by correspondingly increased difficulty for all ages to be tested. If the test curve rises quickly and flattens out, it is not a test of the higher abilities of older children. Since later extension of test items by Heinis requires that scores, and hence steps, be psychologically equal, proof of such equality should be offered. Heinis writes that the equality can in the main be admitted, but offers no evidence for proof. Indeed, such proof is difficult to maintain, and, if the equality cannot be shown, insistence on a procedure which necessitates it will ordinarily bring criticism.

Vermeylen's choice of tests for his intelligence scale may be questioned. A consideration of the tests shows that the majority of them, especially the first half of the series, are testing specific abilities rather than general adaptation to practical situations such as the Binet

scale tests. One might also question Heinis' statement that the length of time required for administering the Vermeylen scale indicates its precision.

Heinis' adaptation of Vermeylen's results through extension in the upper levels to include scores eleven, twelve, and probably thirteen, may be supported on the basis of the trend of the scores for the lower ages and the obvious bunching at the level of score ten. The choice of having approximately forty cases in score ten with thirty to be distributed over eleven and twelve can be followed, but how the median is increased by three and five-tenths points is not obvious, since it is scarcely possible that *each* subject's score would be augmented by that amount inasmuch as the increase is dependent on the existing score. The means will be affected to that extent, but not necessarily the medians.

After calculating the six medians, Heinis fits a logarithmic curve to them presenting only the final equation and none of the mathematical procedure which is sometimes sought by students of the personal constant. To obtain the values for the constants in the equation, he uses P_{11}, the point calculated by theoretical extension to scores of eleven and twelve. Nevertheless, the other three points (P_7, P_9, P_{10}), fall close to the curve, a fact which indicates the acceptability of choosing P_{11}.

Having found the empirical curve between six and twelve years, Heinis proceeds to obtain the parts below and above it through substituting appropriate values of x, that is, age points, in the logarithmic equation. One must bear in mind that this entire curve is based on the results of but sixty children, ten at each of six age levels. Even if these sixty children had been thoroughly examined, it cannot yet be said that another sample of sixty would yield the same results. With statistics based as they are on large frequencies and probabilities, it seems questionable to make predictions on the basis of so few cases.

The fact that the curve fits the median points for years six to twelve does not appear to be sufficient evidence that it will describe the areas below and above those limits. Investigations into other phases of human development indicate the irregularity of growth[3] and suggest the inadvisability of assuming the form of any segment of a growth curve from a preceding or succeeding segment. Heinis presents no evidence here that the theoretical points can be empirically found, although he calculates the number of items necessary for each test in order to provide a complete range of testing.

The application of the personal constant to Binet examination may be questioned, since it would seem to presuppose that mental development as measured by the Binet tests is similar to that measured by the Vermeylen tests. Heinis does not mention in this article the application of the personal constant to the Binet scale. However, he comments at some length in a later part of the article on the lack of precision of metrical scales constructed as the Binet Scale is.

Despite the statistical and theoretical criticisms which may be made of the formulation of the personal constant, its superiority to the IQ for prediction is on the whole supported by empirical evidence. Heinis[5] applied the method of the personal constant to Kuhlmann's[9] results of Binet tests on feeble-minded children, and found that the PC was approximately constant for imbeciles, morons, and borderline subjects. Riley[12] found the mean difference in IQ's between the first and final test over a period of years to be eleven points as contrasted with a mean difference of one point in the corresponding PC's.

Hilden[6] found that the mean discrepancy between actual MA and predicted MA of six hundred thirty-nine feeble-minded subjects on the basis of IQ was as high as −5.91, while the use of the PC for prediction resulted in a mean discrepancy of −1.08. The change in the median IQ increased with the time interval; after one year it was −1.8 and after six years −7.8. The PC varied without trend from −1.0 after two years to +.05 after six years. Hilden[7] reports in another article that Baldwin and Stecher's data converted into IQ and PC curves show the PC to be significantly more constant than the IQ for superior and average children between the ages of eight and twelve.

McIntire's[11] results with forty mentally deficient subjects examined successively over a period of one to ten years are in general agreement with the above studies. Cattell[1] found that the PC was more constant than the IQ for children of sub-average intelligence, but that there was a tendency for the IQ to be slightly more constant than the PC for children of superior intelligence. Hoakley[8] concludes from her findings that "in the extreme cases, mental ages predicted by the PC may vary as much or more than those prediced by the IQ but the extreme difference of the middle fifty per cent of the cases is always smaller than that for the IQ," and that "the PC is a more constant and a more representative index than the IQ."

As Kuhlmann[10] and others have pointed out, the chief value of the PC over the IQ for prediction purposes seems to be that it provides

for the decrease in rate of development of individuals of below average mentality and for the increase in rate of development of those of above average mentality. However, the PC does not escape the criticisms which have been voiced against the IQ. The IQ is constant "on the average" for average individuals, and the personal constant is apparently constant "on the average" for not only average individuals, but for superior and inferior individuals also. For groups and for the "usual" case, such indexes may be used to advantage. But in clinical application our interest is focused on the *individual*, and our main problem consists of making a diagnosis and prognosis for the *unusual individual*. Although growth curves of groups may follow mathematical formulas, the growth of individuals is often quite irregular and unpredictable, and all mathematical indexes may be of little value.

BIBLIOGRAPHY

1. Cattell, P.: "The Heinis Personal Constant as a Substitute for the IQ." *Journal of Educational Psychology*, Vol. XXIV, 1933, pp. 221–228.
2. Descoeurdes, A.: *Le developpement de l'enfant de deux a sept ans.* Neuchatel et Paris, 1921, pp. 16 and 30.
3. Ellis, R. S.: *The Psychology of Individual Differences.* New York: D. Appleton Company, 1928, pp. 217–230.
4. Heinis, H.: "La loi du developpement mental." *Archives de Psychologie*, No. 74, 1924, pp. 97–128.
5. Heinis, H.: "A Personal Constant." *Journal of Educational Psychology*, Vol. XVII, 1926, pp. 163–186.
6. Hilden, A. H.: "A Comparative Study of the Intelligence Quotient and Heinis' Personal Constant." *Journal of Applied Psychology*, Vol. XVII, 1933, pp. 355–375.
7. Hilden, A. H.: *Table of Heinis Personal Constant Values.* Educational Test Bureau, Minneapolis, 1933, pp. 28.
8. Hoakley, Z. P.: "A Comparison of the Personal Constant and the Intelligence Quotient." *Proceedings of the American Association on Mental Deficiency*, Vol. XL, 1935, pp. 403–420.
9. Kuhlmann, F.: "The Results of Repeated Mental Re-examinations of Six Hundred Thirty-nine Feeble-minded over a Period of Ten Years." *Journal of Applied Psychology*, Vol. V, 1921, pp. 195–224.
10. Kuhlmann, F.: "A Median Mental Age Method of Weighting and Scaling Mental Tests." *Journal of Applied Psychology*, Vol. XI, 1927, pp. 181–198.
11. McIntire, J. T.: *The Predictability of the Personal Constant.* The Training School at Vineland, 1934. Unpublished.
12. Riley, G. L.: "A Comparison of the PC and IQ." *Psychological Clinic*, Vol. XVIII, 1930, pp. 261–265.
13. Vermeylen, G.: "Les debiles mentaux (Etude experimentale et clinique)." *Bulletin de l'institute General Psychologie*, No. 4–6, 1922.

A CLASSROOM DEMONSTRATION OF "ATMOSPHERE EFFECT" IN REASONING

SAUL B. SELLS AND HARRY F. KOOB

Columbia University

Few of the reported experiments on thought or reasoning have been found suitable for laboratory courses or classroom demonstrations. Some recent research at Columbia University on the general problem of mental set and directional influences in reasoning[3] has been adapted to classroom use with gratifying success. The present note is the first of several designed to make these procedures generally available. The literature on the problem of mental set in reasoning may be found in the discussions of Boring,[1] Rees and Israel,[2] Sells,[3] and Woodworth and Sells.[4]

"Atmosphere effect" is related to set, Einstellung, attitude, disposition, determining tendency, directive tendency, halo effect, and perseveration. It is a set which arises within an individual, which is derived from and is specific to a particular problem. The individual is not necessarily aware of this set. The "atmosphere effect" occurs in problems or situations in which there is a closed series of responses related to a single task. The result of the "atmosphere effect" is that the individual makes a response (*e.g.*, an inference or judgment) which is most similar in quality to the general trend or tone of the whole situation set up. This is a psychological determiner of an individual's response in a situation.

The following classroom demonstration of "atmosphere effect" is not only a valuable aid to the instructor in presenting the concept of mental set in reasoning, but is highly interesting to the student as well. After the first section, *A*, which presents the procedure for the instructor, the Test Sheet, *B*, the Record Sheet, *C*, and a brief discussion, including some results, *D*, follow.

A. PROCEDURE

1. Distribute the reasoning test (see *B*) to class. Follow routine testing procedure. Students are to comply with directions on page. Answer no questions on test. Time limit: approximately five minutes.

2. Do not collect papers; have students turn over test papers on desks. Distribute Record Sheets (see *C*). Discuss the definition of

"atmosphere effect" printed on Record Sheet, elaborating where necessary to coördinate the material with the course.

3. For further illustration of "atmosphere effect" present several examples, such as the following:

(*a*) Grammatical errors of number: A heedless speaker or writer frequently makes the verb agree in number with the general "atmosphere" of the subject phrase, instead of the noun which is the formal subject*:

> The laboratory equipment in these situations were in many instances essentially the same as those used before.
>
> A common background of emotional conflict, repression, and abnormal reactions of a motor, visceral or perceptual nature have justified an inherent relationship.
>
> Here a series of experiments were conducted with a knowledge of results as the only supposed variables.

(*b*) Word Puzzles: Read the following questions to the class and have the students answer in unison:

(1) What is the correct pronunciation of POLK? (*Ans. poke*, the *l* is silent)

(2) How do you pronounce the word FOLK? (*Ans. foke*, the *l* is silent)

(3) How do you pronounce the white of an egg?

Even if they recognize that the third question refers to the white rather than to the yellow part of an egg, most students will either report or admit a strong tendency to respond with the word YOLK.

4. Make up atmosphere table of class results, by taking the census of responses of each category of conclusions to each set of premises. Types of conclusions may be readily identified by referring to notation described on Record Sheets. Have students record frequencies in appropriate cells of tables on Record Sheet. Place a duplicate table on blackboard.

B. TEST SHEET†

Fill in the blank spaces in the third line of each of the problems below, as in the following example:

---

* The sentences presented here are quoted from the psychological literature.

† The Test Sheet and the Record Sheet should be mimeographed.

Example: All of the students in Psychology are sophomores; And all sophomores are interested in current events; Therefore _____ students in Psychology _____ interested in current events.

The conclusion, correctly filled in, is "Therefore *all* students in Psychology *are* interested in current events."

Work quickly. No questions will be answered until the test is over.

1. All *x*'s are *y*'s;
   And all *y*'s are *z*'s;
   Therefore _____ *x*'s _____ *z*'s.
2. All *x*'s are *y*'s;
   And all *x*'s are *z*'s;
   Therefore _____ *y*'s _____ *z*'s.
3. No *x*'s are *y*'s;
   And no *z*'s are *x*'s;
   Therefore _____ *z*'s _____ *y*'s.
4. Some *x*'s are *y*'s;
   And some *y*'s are *z*'s;
   Therefore _____ *x*'s _____ *z*'s.
5. Some *x*'s are not *y*'s;
   And some *z*'s are not *x*'s;
   Therefore _____ *z*'s _____ *y*'s.

C. RECORD SHEET

Definition and Description of "atmosphere effect."

Definition.—The "atmosphere effect" is a temporary set of the individual, arising within a situation (*e.g.*, problem), to complete a task with that one of several alternative responses (*e.g.*, an inference or judgment) which is most similar to the general trend or tone of the whole situation (*e.g.*, problem).

The "atmosphere effect" is a mental set which arises within the individual and which is derived from and is specific to a particular problem. The individual is not necessarily aware of this set. The "atmosphere effect" occurs in problems or situations in which there is a closed series of responses related to a single task. The result of the "atmosphere effect" is that the individual makes a response (*e.g.*, an inference or judgment) which is most similar in quality to the general trend or tone of the whole situation set up. This is a psychological determiner of an individual's response in a situation.

LOGICAL NOTATION FOR INTERPRETATION OF "ATMOSPHERE EFFECT" IN SYLLOGISTIC REASONING

A—Universal Affirmative.	All *x*'s are *y*'s.	All—yes atmosphere.
E—Universal Negative.	No *x*'s are *y*'s.	All—no atmosphere.
I—Particular Affirmative.	Some *x*'s are *y*'s.	Some—yes atmosphere.
O—Particular Negative.	Some *x*'s are not *y*'s.	Some—no atmosphere.

TABLE FOR RECORDING CLASS RESULTS

Problem	Premises	Conclusions				
		A	*E*	*I*	*O*	None
1	*AA*					
2	*AA*					
3	*EE*					
4	*II*					
5	*OO*					

Number of students in class_____

D. DISCUSSION

Tables I and II show the results obtained with one class of twenty-three students, and the combined results of five classes, totalling one hundred thirty-four students at Columbia College.

TABLE I.—RESPONSES OF TWENTY-THREE STUDENTS TO TEST
Entries Indicate the Number of Students Offering Each Type of Conclusion, or No Conclusion, to Each Set of Premises

Problem	Premises	Conclusions				
		A	*E*	*I*	*O*	None
1	*AA*	23	0	0	0	0
2	*AA*	16	1	6	0	0
3	*EE*	0	15	1	0	7
4	*II*	0	0	19	0	4
5	*OO*	0	0	3	17	3

The significance of these tables has been discussed at length in a monograph[3] which is readily accessible. It may be readily seen that the results demonstrate "atmosphere effect," since in each case the greatest frequency falls in the cell of the table where the conclusion is most similiar in quality to the premises from which it is drawn. With reference to the five problems in the test, it may be pointed out that both a valid *A* and a valid *I* conclusion are possible in problem one, only a valid *I* conclusion in problem two, and no possible valid conclu-

TABLE II.—RESPONSES OF ONE HUNDRED THIRTY-FOUR STUDENTS TO TEST Entries Indicate the Number of Students Offering Each Type of Conclusion, or No Conclusion, to Each Set of Premises

Problem	Premises	Conclusions				
		A	*E*	*I*	*O*	None
1	*AA*	133	0	1	0	0
2	*AA*	104	1	26	0	3
3	*EE*	3	88	23	4	13
4	*II*	0	0	121	1	12
5	*OO*	0	3	18	94	19

sions in problems three, four, and five. The correct conclusions in problem 1 are "All x's are z's" and "Some y's are z's." The correct conclusion in problem two is "Some y's are z's."

This demonstration has been timed at approximately twenty-five minutes in nine classes at Columbia College. The nature of the test material, syllogisms, upon which this demonstration of "atmosphere effect" is based, restricts its use to students who have not had courses in logic, although not entirely, since forgetting of this type of material is frequently rapid. It is suggested that if there are any logic students in the class, the instructor treat these as a separate group.

BIBLIOGRAPHY

1. Boring, E. G.: *A History of Experimental Psychology.* New York: The Century Company, 1929, pp. xv + 699.
2. Rees, H. J. and Israel, H. E.: "An Investigation of the Establishment and Operation of Mental Sets," in Psychol. Studies from Smith College. Gibson, J. J., ed. *Psych. Mon.* 210, 1935.
3. Sells, S. B.: "The Atmosphere Effect: An Experimental Study of Reasoning." *Arch. of Psych.*, 200, 1936.
4. Woodworth, R. S. and Sells, S. B.: "An Atmosphere Effect in Formal Syllogistic Reasoning." *J. Exp. Psych.*, Vol. XVIII, 1935, pp. 451–460.

A COMPARISON OF THE PERFORMANCE OF MATCHED GROUPS OF SCHIZOPHRENIC PATIENTS, NORMAL SUBJECTS, AND DELINQUENT SUBJECTS ON SOME ASPECTS OF THE STANFORD-BINET

CHARLOTTE HALL ALTMAN AND DAVID SHAKOW*

Worcester State Hospital, Worcester, Mass.

Wells and Kelley,[1] and Wentworth,[2] among others, have analyzed the successes and failures of psychotic patients on the Stanford-Binet examination. As far as can be determined, no attempt has been made, however, to match patients with normal subjects for mental age, and then to compare the performance of these "matched pairs" on the different sub-tests of the scale. In the present study, matched groups of schizophrenic patients, normal controls, and adult delinquents were compared in this way. It was found, when the data were scrutinized, that the patients tended to surpass their matched normal subjects in digit-repetition at the VII- (digits forward), XII- (digits backward), and XIV- (digits forward) year levels, in the induction test (XIV-2), the arithmetical reasoning test (XIV-5), picture interpretation (XII-7), reproduction of designs (X-3) and the code test (XVI-6). Their most marked relative success, however, was in the vocabulary test; they excelled in this at every age-level in which the test is given, showing marked superiority at the XII- and XVI-year levels. On the other hand, patients failed the absurdities test (X-2) much more frequently than the normals. Fable interpretation (XII-5 and XVI-2) and the Ball-and-Field test (VIII-1 and XII-3) likewise proved more difficult for patients than for either of the other groups. The deviation of the patients from the other subjects was, however, most pronounced in the vocabulary test. This is in accordance with the results of Wells and Kelley[1] and Dearborn.[3] Babcock,[4] on the basis of a similar finding, devised an efficiency index. Using the vocabulary test as the measure of pre-psychotic ability, she compared vocabulary level with the level achieved on a battery of tests emphasizing speed and efficiency of learning. She reports on the use of the test with schizophrenic patients in a later publication.[5] Schwartz,[6] Wittman[7] and Simmons[8,9] are others who have made use of Babcock's device or of a similar technique in testing psychotic patients. All of these workers report that differ-

* From the Research Service of the Worcester State Hospital.

ences were found, always in the direction of superior vocabulary performance as contrasted with performance on other tests. A survey of the present material from the point of view of the discrepancy between vocabulary scores and scores on the other parts of the Stanford-Binet was therefore undertaken under conditions where the important factor of mental level was controlled.

SUBJECTS

The cases selected for this survey were all male, including schizophrenic patients, adult delinquents, tested routinely under the Briggs law (recidivists, or first offenders accused of capital crimes),[10] and normal controls, who were chiefly hospital employees and men especially hired as controls for a series of physiological and psychological tests. The Stanford-Binet intelligence scale was used. The coöperation of the subjects was graded in each case by the examiner, a grade of *A* or *B* indicating that, in differing degrees, the subject was actively interested in the tests and apparently exerted effort to do his best; a grade of *C* or *D* indicating that, in differing degrees, the subject was either

TABLE I.—CONSTANTS OF THE DISTRIBUTION OF CHRONOLOGICAL AGE AND EDUCATION OF FIFTY-SIX MATCHED SCHIZOPHRENICS, NORMAL SUBJECTS, AND DELINQUENT SUBJECTS GIVING REPRESENTATIVE RESULTS AND THIRTY MATCHED SUBJECTS GIVING NON-REPRESENTATIVE RESULTS

	Chronological age in years					Education-grade attained				
Group	No.	Minimum	Maximum	Mean	SD	No.	Minimum	Maximum	Mean	SD
Representative results (*A-B* coöperation)						Representative results				
Schizophrenics	56	14	46	29.1	7.4	56	2	17	8.4	2.7
Normal subjects	56	17	47	26.1	7.1	56	4	14	9.2	2.3
Delinquents	56	17	56	24.1	7.0	56	3	17	7.7	4.0
Non-representative results (*C-D* coöperation)						Non-representative results				
Schizophrenics	30	18	51	27.3	7.7	30	4	16	9.2	2.8
Normal subjects[1]	30	17	37	26.6	6.4	30	4	12	8.5	2.0
Delinquents	30	16	56	26.5	9.0	30	2	10	6.9	1.6

[1] Since no normal subjects of *C-D* coöperation level were available, those with *A-B* level were used throughout for comparison.

apathetic and indifferent towards the test or antagonistic, exerting only perfunctory or spasmodic effort. All subjects selected were sufficiently coöperative to make complete results on the examination possible, however. The subjects were divided into two groups, an *A-B* coöperation group and a *C-D* coöperation group. Test results on the *A-B* group were considered representative of the subject's actual ability at the time of the administration of the test; results on the *C-D* group were not representative because of the poor coöperation given. In each of these groups, patients, normal subjects, and delinquents were matched individually for mental age on the complete Stanford. Since all of the normal subjects had coöperation grades of *A* or *B*, it was necessary to match some patients and delinquents of *C-D* coöperation with normals of *A-B* coöperation. There were fifty-six subjects each in the patient, normal, and delinquent groups, matched for mental age and *A-B* coöperation; and thirty cases in each group with *C-D* coöperation. Although no attempt was made to match the cases for chronological age and education, except to keep the whole group within the same general range, scrutiny of Table I will show that the groups did not differ widely in these respects.

TABLE II.—CONSTANTS OF THE DISTRIBUTION OF COMPLETE STANFORD-BINET MENTAL AGES IN MONTHS TOGETHER WITH INTERCORRELATIONS, OF MATCHED GROUPS OF SCHIZOPHRENICS, NORMAL SUBJECTS, AND DELINQUENT SUBJECTS GIVING REPRESENTATIVE RESULTS, AND OF MATCHED GROUPS GIVING NON-REPRESENTATIVE RESULTS

Group	No.	Minimum	Maximum	Mean	SD	Intercorrelations
Mental age in months—complete Stanford-Binet						
Representative Results						
Schizophrenics	56	112	228	157.1	25.9	Patients *vs.* normals + .9993
Normal subjects	56	111	228	157.0	25.9	Patients *vs.* delinquents + .9987
Delinquents	56	111	222	157.0	25.6	Normals *vs.* delinquents + .9985
Non-representative Results						
Schizophrenics	30	109	177	140.9	17.7	Patients *vs.* normals + .9986
Normal subjects	30	111	176	141.0	17.5	Patients *vs.* delinquents + .9864
Delinquents	30	108	177	141.1	16.9	Normals *vs.* delinquents + .9858

The success of the matching for Stanford mental age may be judged from Table II; in no case did the members of a matched trio differ from each other by more than six months of mental age; in most cases the difference was not greater than one month.

The psychiatric sub-types of the patients in the two groups were as follows: In the "Representative" group, hebephrenic, three cases; catatonic, seven; simple, five; paranoid, seventeen; mixed, four; late indeterminate, two; unclassified, eighteen. In the "Non-representative" group, hebephrenic, seven; catatonic, ten; simple, two; paranoid, two; late indeterminate, one; unclassified, eight.

TECHNIQUE AND RESULTS

In order to obtain a "discrepancy score" which would show the difference between the level attained on the vocabulary test alone, and performance on the remainder of the Stanford scale, the following procedure was used: The vocabulary scores were distributed for the "Representative" and the "Non-representative" groups separately, the deviation from the mean of the group in terms of standard deviation was computed for each subject, and may be called the standard deviation, or σ-score on the vocabulary test. The Stanford examination of each subject was then re-scored, omitting the vocabulary test and giving proportional credit to the other sub-tests at the age-levels where the vocabulary test occurs. These mental ages on the Stanford without the vocabulary test were similarly distributed, and σ-scores obtained for the "Representative" and for the "Non-representative" groups separately. The σ-scores in the "Representative" group ranged from -2.505 to $+2.667$ on the vocabulary test, and from -2.268 to $+2.633$ on the Stanford without vocabulary. In the "Non-representative" group, the σ-scores ranged from -2.023 to $+2.784$ on the vocabulary test, and from -2.403 to $+2.136$ on the Stanford without vocabulary.* The difference between the σ-score on the vocabulary test and the σ-score on the Stanford without the vocabulary was then considered to represent the discrepancy in per-

* The mean absolute scores for vocabulary were: Representative group—Schizophrenics, 54.9; Normals, 49.3; Delinquents, 51.4; Non-representative group—Schizophrenics, 52.0; Normals, 40.8; Delinquents, 42.0. The mean absolute values in months for Stanford-Binet without vocabulary were: Representative group—Schizophrenics, 156.0; Normals, 158.0; Delinquents, 157.4; Non-representative group—Schizophrenics, 137.5; Normals, 142.0; Delinquents, 141.7.

formance on vocabulary as opposed to other tests of the Stanford, and a discrepancy score thus obtained for each subject. The constants of the distributions of the discrepancy scores are given in Table III for patients, normal subjects, and delinquents separately, a plus score indicating a discrepancy in favor of the vocabulary.

TABLE III.—CONSTANTS OF THE DISTRIBUTIONS OF DISCREPANCY SCORES FOR BOTH REPRESENTATIVE AND NON-RERPESENTATIVE MATCHED GROUPS OF SCHIZOPHRENICS, NORMAL SUBJECTS, AND DELINQUENT SUBJECTS

Group	No.	Discrepancy scores: Minimum	Maximum	[1]Mean and SE_M	SD and SE_{SD}
Representative Results					
Schizophrenics	56	−1.786	+1.927	+0.265 ± 0.101	0.750 ± 0.071
Normal subjects	56	−1.639	+1.739	−0.220 ± 0.099	0.735 ± 0.070
Delinquents	56	−1.767	+1.663	−0.045 ± 0.102	0.758 ± 0.072
Non-representative Results					
Schizophrenics	30	−1.404	+3.012	+0.706 ± 0.208	1.121 ± 0.147
Normal subjects	30	−2.088	+2.183	−0.398 ± 0.158	0.851 ± 0.112
Delinquents	30	−2.526	+1.098	−0.292 ± 0.169	0.908 ± 0.119

[1] Plus scores indicate discrepancy in favor of vocabulary; minus scores in favor of the rest of the Stanford.

In both the "Representative" and the "Non-representative" groups, the distributions show noteworthy differences among the three types of subjects used; these differences are particularly marked in the distributions of the "Non-representative" group. In each case, the discrepancy scores of the normal subjects and delinquents overlap markedly; their means are all on the minus side, though the delinquents are closer to zero than the normal subjects (*i.e.*, they show somewhat less discrepancy between scores). The mean discrepancy scores of the schizophrenics, on the other hand, are both plus, indicating that their performance on the vocabulary test is relatively better than that of the other groups when compared with performance on the rest of the Stanford.

In order to test the significance of the difference between these mean discrepancy scores and zero, Fisher's *t* test was used. The same test

was used to determine the significance of the difference between the mean scores of schizophrenics, normal subjects, and delinquents. The *t*'s, with their corresponding probabilities, are given in Table IV.

TABLE IV.—SIGNIFICANCE OF DISCREPANCY SCORES OBTAINED IN BOTH REPRESENTATIVE AND NON-REPRESENTATIVE MATCHED GROUPS OF SCHIZOPHRENICS, NORMAL SUBJECTS, AND DELINQUENT SUBJECTS

Fisher's *t*—significance of discrepancy scores

Significance of difference between mean score and zero			Significance of differences between mean scores		
Group	*t*	*P*	Group	*t*	*P*
Representative Results					
Schizophrenics	2.62	.01	Patients *vs.* normal controls	3.28	< .002
Normal subjects	2.22	.03	Patients *vs.* jail cases	2.15	.03
Delinquents	0.44	> .5	Normals *vs.* jail cases	1.23	.23
Non-representative Results					
Schizophrenics	3.39	.002	Patients *vs.* normal controls	3.92	.0002
Normal subjects	2.52	.018	Patients *vs.* jail cases	3.58	.0006
Delinquents	1.73	.100	Normals *vs.* jail cases	0.44	.7 > *P* > .6

In both the "Representative" and the "Non-representative" groups, then, the mean discrepancy scores are significantly different from zero for the schizophrenics and the normal subjects, but not for the delinquents. In both groups, too, the schizophrenics differ significantly from the normal subjects and from the delinquents. The normal and delinquent groups do not differ significantly from each other at either coöperation level, however. Neyman and Pearson's λ-test[11] was applied to determine whether any two of the distributions of samples could have been drawn from a common population with unspecified mean and standard deviation. The previous results were corroborated in each instance; again the schizophrenics differ significantly from both the normals and the delinquents, while the difference between normals and delinquents is not significant. The schizophrenics were then tested against the normals and delinquents merged into one distribution. The λ's with their approximate *P*'s are given in Table V.

TABLE V.—NEYMAN AND PEARSON'S λ-SIGNIFICANCE OF DIFFERENCE BETWEEN DISTRIBUTIONS OF DISCREPANCY SCORES OBTAINED IN BOTH REPRESENTATIVE AND NON-REPRESENTATIVE MATCHED GROUPS OF SCHIZOPHRENICS, NORMAL SUBJECTS, AND DELINQUENT SUBJECTS

Group	λ	*P*
Representative Results		
Schizophrenics *vs.* normals	.0034	$P < .01$
Schizophrenics *vs.* delinquents	.0327	$.01 < P < .05$
Normals *vs.* delinquents	.4539	$P > > .05$
Schizophrenics *vs.* merged normals and delinquents	.0034	$P < < .01$
Non-representative Results		
Schizophrenics *vs.* normals	.0001	$P < .01$
Schizophrenics *vs.* delinquents	.0008	$P < .01$
Normals *vs.* delinquents	.8433	$P > > .05$
Schizophrenics *vs.* merged normals and delinquents	.000008	$P < < .01$

Since the λ-method tests the joint significance of the difference between both mean and standard deviation, thus extending the significance of the differences between means found by Fisher's *t* method, it would seem that the findings may be accepted with considerable confidence.

Having established a reliable difference between the discrepancy scores of schizophrenics and those of normal subjects and delinquents, some explanation of this difference in performance was sought. The greater discrepancy found in "non-representative" groups is an interesting, and quite significant, finding. It indicates that even when coöperation on the test is poor, the vocabulary test is affected relatively less, perhaps because it receives relatively greater attention and effort. This suggests that a vocabulary test, standardized as a separate test, might yield a better estimate of mental level than the complete Stanford Examination in less coöperative psychotic patients. However, variation in coöperativeness on the various sub-tests of the Stanford, if present, can account for only part of the differences obtained. The delinquents of the same poor coöperation level did not show the superiority on the vocabulary test which was found in the schizophrenics; they, indeed, obtained a discrepancy score much like that of the normal subjects with whom they were matched, but whose results

were considered representative. Furthermore, in the schizophrenics, normals, and delinquents who give representative results, similar differences obtain, though they are somewhat less marked.

In view of the results of Babcock and the others who have written on the subject (Schwartz, Wittman, and Simmons), it was at first believed that the discrepancy scores might represent the same kind of phenomenon which they have accepted as a measure of deterioration. If the discrepancy score is a valid measure of the difference between performance on a vocabulary test and on other tests, this would seem to be a reasonable assumption, even though tests "which emphasize new learning and speed" (Babcock) were not used except as they are contained in the Stanford. To determine the validity of this assumption, various characteristics of the patient group were taken for comparison. Period of hospitalization was tentatively considered as a possible criterion of deterioration, although its limitations as such a criterion were recognized. The coefficients of correlation between period of hospitalization and discrepancy score was +.09 for the schizophrenics giving representative results, and +.02 for the schizophrenics giving non-representative results.

A more satisfactory criterion was sought in estimates of intellectual and of emotional deterioration submitted by the psychiatrists in charge of the patients. These estimates consisted of grades on a five-point scale, ranging from total absence of deterioration to very marked deterioration, made routinely together with grades on thirty-six other traits. These were made before the present study was begun, so that no question of bias on the part of the psychiatrists in making their ratings can arise. The results show entire lack of association between the estimates of deterioration and the discrepancy scores. This was as marked in the case of ratings of intellectual as of emotional deterioration. There is no evidence, therefore, insofar as the few measures used are concerned, that the discrepancy score is a measure of deterioration.

Further clues as to the meaning of the discrepancy scores were sought in correlations of these scores with education and with chronological age. Again there is no evidence of any direct association between the discrepancy scores and either of these variables; for schizophrenics, normal subjects, and delinquents, for representative as well as non-representative results, the correlations are all close to zero. It was further postulated that, if mental age on the complete Stanford is taken to represent present intellectual level, the degree of associa-

tion between present level and the discrepancy score might be informative. Correlation coefficients were, therefore, computed for these two variables. They range from −.10 to −.50, both extremes coming from the normal group, the former from the cases matched with schizophrenics and delinquents giving representative results, the latter from the cases matched with schizophrenics and delinquents giving non-representative results. When one outlying case is omitted, even the correlation of −.50 drops to −.41. Variation in present intellectual level, therefore, accounts for very little—at most 14 per cent—of the variation in discrepancy score. It should be noted, though, that there is some spuriosity in these correlations, since the discrepancy score is probably some function of the mental age on the complete Stanford. It seems reasonable to assume that the true correlation between present intellectual level and discrepancy score would be even lower than those reported above. The correlations, however, are all in the "expected" direction, *i.e.*, the slight trend which does occur indicates that the subjects with the lowest mental ages show relatively better vocabulary performance than other subjects, and *vice versa*.

DISCUSSION

That there is a difference in the performance of psychotic and non-psychotic subjects on the vocabulary test as opposed to the other tests of the Stanford scale, seems to be definitely established. Our data, however, have adduced no evidence that the discrepancy in performance represents intellectual or emotional deterioration. In their reports, Babcock and Schwartz both indicate a relationship between the degree of deterioration and the efficiency index, but neither gives any correlation coefficients. Babcock distributes her cases according to efficiency index, then discusses the increase in deterioration in successive arrays of increasing "inefficiency." Schwartz presents a graph showing the association between period of hospitalization and efficiency index, but it is evidently a graph of the means of the arrays only. It is not possible to determine the extent of the scatter around these means, but from the great irregularity of the graph one would judge that the scatter is large (*e.g.* the mean efficiency index is the same for patients hospitalized for two and for fourteen years). The relationship is clearly present, even if it is a loose one, however. It may well be that the failure of the present findings to substantiate those of Babcock and Schwartz is due to the fact that the discrepancy score is not strictly comparable to Babcock's efficiency index. In any case, it seems clear that the

obtained discrepancy scores do not measure deterioration as it is at present clinically defined.

On the other hand, it seems equally evident that relatively high discrepancy scores are a distinguishing feature of schizophrenic patients. Further investigation of the problem is needed and might well give clues leading to a better understanding of schizophrenic thought processes. The kinds of tests on which schizophrenics differ markedly in performance from normal people of the same mental age may lead to some elucidation of the kinds of mental activity which characterizes them. When the differences can be demonstrated to be statistically reliable, as in the present case, when better criteria of psychiatric concepts, such as deterioration, are available, and when results of a number of similar and related investigations are compared and integrated, studies such as the present offer hope that they may form the basis for an objective measure of some aspects of schizophrenic mental activity.

SUMMARY

The object of the present inquiry was to compare the vocabulary level with the general intellectual level in psychotic and non-psychotic subjects. Two groups of schizophrenic patients, normal subjects, and adult delinquents were matched for mental age on the Stanford. The first group included subjects of *A* or *B* coöperation level whose test results were considered representative of the subject's ability at the time of the test; the second, those of *C* or *D* coöperation, whose results were not considered representative. The Stanford vocabulary scores of each group were distributed, and the score of each individual was expressed in terms of deviation from the mean in standard-deviational units. Similarly, σ-scores on the remainder of the Stanford, with the vocabulary test omitted, were obtained. The difference between these σ-scores, termed the "*discrepancy score*," was considered to represent the relative performance on the vocabulary test, as compared to performance on the rest of the Stanford. In both coöperation groups, schizophrenics were found to have higher discrepancy scores (indicating relatively better vocabulary ratings) than either normal subjects or delinquents. The differences were found to be statistically significant in both instances. No association between discrepancy scores and three different criteria of deterioration could be established. The scores, then, probably do not represent degree of deterioration in the schizophrenics, as Babcock's "efficiency index" is presumed to do. Further investigation may show that the discrepancy scores are a

measure of some aspect of "thinking disturbance" in schizophrenics, related or unrelated to deterioration.

BIBLIOGRAPHY

1. Wells, F. L. and Kelley, C. M.: "Intelligence and psychosis." *Amer. J. Insan.*, Vol. LXXVII, 1920–1921, pp. 17–45.
2. Wentworth, M. M.: "Two hundred cases of dementia praecox tested by the Stanford Revision." *J. Abn. Psychol.*, Vol. XVIII, 1923–1924, pp. 378–384.
3. Dearborn, G. V. N.: "The determination of intellectual regression and progression." *Amer. J. Psychiat.*, Vol. VI, 1926–1927, pp. 725–741.
4. Babcock, H.: "An experiment in the measure of mental deterioration." *Arch. Psychol.*, Vol. XVIII, No. 117, 1930, p. 105.
5. Babcock, H.: *Dementia Praecox, a psychological study.* Lancaster, Pa.: Science Press Printing Co., 1933, p. 167.
6. Schwartz, L.: "Measurement of mental deterioration in Dementia Praecox." *Amer. J. Psychiat.*, Vol. XII, 1932–1933, pp. 555–562.
7. Wittman, P.: "The Babcock Deterioration Test in State Hospital practice." *J. Abn. & Soc. Psychol.*, Vol. XXVIII, 1933–1934, pp. 70–84.
8. Simmons, C.: "Studies in experimental psychiatry. IV. Deterioration of 'g' in psychotic patients." *J. Ment. Sci.*, Vol. LXXIX, 1933, pp. 704–734.
9. Simmons, C.: "Mental incapacity: the intelligence of patients in mental hospitals." *Charac. & Person.*, Vol. IV, 1935, pp. 25–33.
10. Shakow, D. and Millard, M. S.: "A psychometric study of one hundred fifty adult delinquents." *J. Soc. Psychol.*, Vol. VI, 1935, pp. 437–457.
11. Neyman, J. and Pearson, E. S.: "On the problem of two samples." *Bull. Internat. l'Acad. Pol. des Sci. et Lett.*, Series A, 1930, pp. 73–96.

CERTAIN ASPECTS OF THE VALIDITY OF THE BERNREUTER PERSONALITY INVENTORY

WALTER F. ST. CLAIR AND J. CONRAD SEEGERS

Temple University

Because of the emphasis lately placed upon personality analysis both in educational institutions and elsewhere, and because of the frequent use of the Bernreuter Personality Inventory in this work, considerable interest has been manifested in the Inventory, which has been attacked and defended in a number of recent publications.

The Bernreuter Personality Inventory attempts to measure four separate and distinct aspects of personality, or "traits," at one time. It is this rather unique characteristic which recommends it to certain individuals but which has been attacked most vigorously by others. The separate sections or instruments of the Inventory, it is claimed, measure, separately and simultaneously, "Neurotic Tendency," "Self-Sufficiency," "Introversion-Extroversion," and "Dominance-Submission." On the basis of intercorrelation of the test items and correlations between the Inventory and certain other published instruments which purported to measure specific traits, Bernreuter has stated that there seems reason to believe that definite and distinct personality traits, which exist as separate entities, are measured by the Inventory.[1] More recently[2] he has suggested that possibly only three distinct traits emerge as the result of applying the inventory, stating that the measures of "Neurotic Tendency" and of "Introversion" cannot be clearly distinguished.

The attacks upon the Bernreuter Personality Inventory may be illustrated by the contentions of Lorge and of Flanagan. Lorge has criticized[3] the validation of the Inventory through intercorrelation of the test items, stating that his analyses indicate that the "correlations in the scores are attributable to the weighting system" inherent in the Inventory itself. He concludes that "probably not more than two independent traits can be used to account for the 'traits' of the

[1] Bernreuter, Robert G.: *Manual for Personality Inventory.* Stanford University Press, 1935.

[2] Bernreuter, Robert G.: "The Imbrication of Tests of Introversion-Extroversion and Neurotic Tendency." *Journal of Social Psychology,* 1934, pp. 184–199.

[3] Lorge, Irving: "Personality Traits by Fiat I. The Analysis of the Total Trait Scores and Keys of the Bernreuter Personality Inventory." *Journal of Educational Psychology,* Vol. XXVI, April, 1935, pp. 273–278.

Bernreuter." Certainly, he insists, there are less than four distinct traits involved.

In another publication, Lorge, Bernholz, and Sells[1] examine the interrelationship of "*Yes*," "*No*," and "?" responses by means of correlation technique. The article states that the low correlations discovered, taken with other evidence cited or presented, indicate that the Bernreuter Personality Inventory lacks reliability, consistency, and independence of trait measurement. A correction[2] withdrew this conclusion as it applied to the *F*1-*C* and *F*2-*S* scales of Flanagan.

Flanagan's[3] approach was by means of the multiple factor analysis. Briefly, he concluded that the scores on the four scales may be largely accounted for in terms of two factors.

In a recent article,[4] Bernreuter presented evidence to the effect that what he calls a "line of preference" is clearly indicated in the answers obtained from college students to whom his Inventory was administered, that more than chance was operant in the results obtained, that evidently something was being definitely measured by the separate measures of the Inventory.

It was partly because of this controversy, partly to try to discover what evidence something over one thousand Bernreuter scores provided a personnel office, that the present study was conducted.

This paper reports some findings derived from the administration of the Bernreuter test and certain other instruments to seven hundred twenty-nine men and four hundred thirty-three women freshmen in Temple University upon entrance, during the scholastic years 1934–1935. This number included most of the entering freshmen. In addition to the Bernreuter a group of questionnaires was administered in an attempt to elicit information from these freshmen concerning their health, home conditions, personal problems of various natures, previous interests and activities, and certain social attitudes.

---

[1] Lorge, I., Bernholz, E., and Sells, S. B.: "Personality Traits by Fiat II. The Consistency of the Bernreuter Personality Inventory by the Bernreuter and Flanagan Keys." *Journal of Educational Psychology*, Vol. XXVI, September, 1935, pp. 427–434.

[2] Lorge, Irving: "Personality Traits by Fiat II. A Correction." *Journal of Educational Psychology*, Vol. XXVI, December, 1935, pp. 652–654.

[3] Flanagan, John C.: *Factor Analysis in the Study of Personality.* Stanford University Press, 1935, pp. 103.

[4] Bernreuter, R. G.: "Chance and Personality Inventory Scores." *Journal of Educational Psychology*, Vol. XXVI, April, 1935, pp. 279–283.

This information was gathered through the offices of the Deans of Women and Men and emphasis was placed upon the thought that the motive was to assist freshmen in adjusting themselves. Every effort was made to assure the students that the information would be kept confidential and to encourage frank and honest responses. Subsequent personal conferences with some forty per cent of the population of the study indicated that the responses were genuine.

A percentile distribution of raw scores with reference to each Bernreuter trait was computed to determine the applicability of Bernreuter's norms to this group. The difference between the two percentile distributions was noticeable, particularly in *B3-I* and *B4-D*.

TABLE I.—PERCENTILE NORMS FOR BERNREUTER PERSONALITY INVENTORY AND NORMS ESTABLISHED FOR TEMPLE FRESHMEN

	Raw scores	Bernreuter college men	Freshman men 729	Bernreuter college women	Freshman women 433		Raw scores	Bernreuter college men	Freshman men 729	Bernreuter college women	Freshman women 433
	125	96	99	98	99		75	97	98	97	98
	100	95	97	96	97		50	92	94	89	95
	75	93	95	93	93		25	83	88	79	85
	50	90	91	88	88	*B3-I*	0	70	77	62	69
	25	83	86	80	80		−25	53	62	42	52
	0	76	79	71	70		−50	35	43	24	33
B1-N	−25	68	70	59	59		−75	17	24	11	13
	−50	58	62	45	46		−100	4	7	3	3
	−75	48	49	35	36						
	−100	34	37	23	24		150	97	97	99	99
	−125	22	24	15	12		125	90	90	94	97
	−150	12	12	6	6		100	80	77	86	90
	−175	6	5	2	1		75	67	62	73	78
							50	52	47	60	62
	125	96	98	98	99	*B4-D*	25	37	30	45	46
	100	91	93	94	97		0	25	20	33	31
	75	83	85	89	92		−25	13	10	20	18
	50	68	71	79	82		−50	8	7	10	10
B2-S	25	50	53	64	67		−75	4	4	5	4
	0	31	33	45	49		−100	3	2	1	2
	−25	16	17	28	33						
	−50	8	6	15	16						
	−75	3	2	6	7						
	−100			2	3						

Although there is sufficient evidence presented here to indicate that norms should be established for different levels of maturity, the differences are not sufficiently significant to affect this study greatly. Consequently, the published norms were employed in the tabulations following. The actual difference is indicated in Table I.

In the questionnaire responses certain situations developed which might lead one to expect abnormal personality traits or responses. The following paragraphs analyze the Bernreuter scores of the several groups concerned in these situations. Bernreuter scores are expressed in percentiles. Accordingly, in the case of a normal group, one would expect a median and a mean of fifty per cent, and one would assume that deviations from fifty per cent indicated some degree of abnormality. On the same assumption one would expect deviations in the scores of a group, when conditions are apparently conducive to abnormality, if the test measures personality traits.

In the analysis the following procedure is employed. In each instance the arithmetic mean and median Bernreuter scores are displayed for the group under discussion. In inspecting these figures it should be kept in mind that a normal group should have a mean of fifty per cent.

Table II displays Bernreuter scores of one hundred students who in the questionnaire reported serious disagreement in their homes as indicated. In this table as in the following tables, means are stated in terms of the nearest whole number. In each of the tables in this paper column *B*1 lists scores in neurotic tendency, *B*2 scores in self-sufficiency, *B*3 scores in introversion-extroversion, *B*4 scores in dominance-submission.

According to Bernreuter's interpretation, a score above 50 in *B*1 indicates neurotic tendency; a score above 50 in *B*2 indicates either independence or withdrawal, a score below 50 dependence; a score above 50 in *B*3 indicates introversion, a score below 50 extroversion; a score above 50 in *B*4 indicates dominance, below 50 submission.

In Table II in every case except that of disagreement between sister and self, the mean in *B*1 is higher than normal. Every median, including that of the group mentioned, is considerably higher than normal. This indicates significant neurotic tendency in each of the groups. The greatest deviation is indicated in the groups reporting disagreement between mother and self, brother and self, and brother and sister.

TABLE II.—SERIOUS DISAGREEMENT IN HOME

Disagreement between	Number of cases		*B*1	*B*2	*B*3	*B*4
Father and self	15	Mean	60	61	61	56
		Median	58	75	58	58
Mother and self	10	Mean	67	66	60	39
		Median	80	67	55	35
Mother and father	30	Mean	55	51	52	49
		Median	56	50	52	53
Brother and self	15	Mean	67	30	64	32
		Median	73	23	65	33
Sister and self	9	Mean	49	42	53	55
		Median	63	35	55	65
Brother and sister	6	Mean	68	38	65	25
		Median	80	35	70	10
Others	15	Mean	52	51	51	54
		Median	52	44	55	48
Total group	100	Mean	57	49	57	47
		Median	59	46	56	45

The *B*2 scores are above normal in some instances, below or normal in others. This score has never been interpreted exactly,[1] although Bernreuter has offered an interpretation referred to above. The figures in the table indicate no definite trend toward either self-sufficiency or dependence.

Both means and medians of *B*3 are consistently above the norms. The highest scores are in cases of disagreement between brother and self, and between brother and sister. The fact that the last named is the highest is interesting, because the individual himself is not directly concerned. It might reflect the effect of a discordant environment upon a sensitive nature.

The actual distribution of most of the *B*4 scores is not indicated by the mean. The distributions were in most instances highly irregular. Some were concave, with cases at the extremes. Most were multi-modal. Only in the cases of disagreement between mother and father, sister and self, brother and sister were the distributions

[1] Other evidence derived from the analyses of these scores indicates as a possible basis of interpretation the relation of *B*2 and the other scores of the inventory. The data, however, are not yet complete. It seems possible that high *B*2 scores accompanied by high *B*1 and *B*3 indicate withdrawal.

fairly indicated by the average. However, in the whole group of scores, lack of dominance is indicated.

In general, it should be noted from Table II that students reporting disagreement in the home are indicated to be inclined toward neurotic tendency and introversion by the Bernreuter Inventory. However, it should be noted also that in each instance the small numbers minimize the significance of these figures.

TABLE III.—STUDENTS DESIRING INTERVIEWS

Problems	Number of cases		*B*1	*B*2	*B*3	*B*4
Health	44	Mean	59	49	57	48
		Median	63	46	60	47
Family relations	5	Mean	73	55	73	45
		Median	75	55	83	48
Religious	25	Mean	61	55	63	47
		Median	71	65	67	45
Social	66	Mean	64	47	58	45
		Median	64	44	63	43
Sex	19	Mean	58	55	60	52
		Median	72	68	65	58
Financial	304	Mean	51	51	50	51
		Median	51	50	50	52
Total group	463	Mean	54	51	53	49
		Median	56	49	52	49

Normally one should expect persons desiring interviews to manifest at least slightly neurotic tendencies, the degree varying with the nature of the problems. In Table III there is a consistent trend in this expected direction, except in the cases of students having financial problems. The means for this group are at the normal level. But this also is what one should expect. Financial difficulties make no personal distinctions. The trends indicated are toward neurotic tendency and introversion.

Again it is to be noted that the *B*2 scores present no clear-cut picture. The *B*4 scores are not truly indicative. The reason is indicated by the *B*4 section of Diagram 1. The mean is normal because the frequencies are heavy at the extremes, but the average deviation of individuals from the norm is considerable. This kind of distribution was typical of the *B*4 and *B*2 groups. It was not typical of *B*1 and *B*3.

In Table IV note the neurotic tendency indicated by the *B*1 scores of individuals who have had heart trouble or Infantile paralysis. Note also that in these groups the *B*4 scores indicate lack of dominance and the *B*2 scores are high. This seems to lend some credence to the suggestion previously advanced that high *B*2 scores, accompanied by

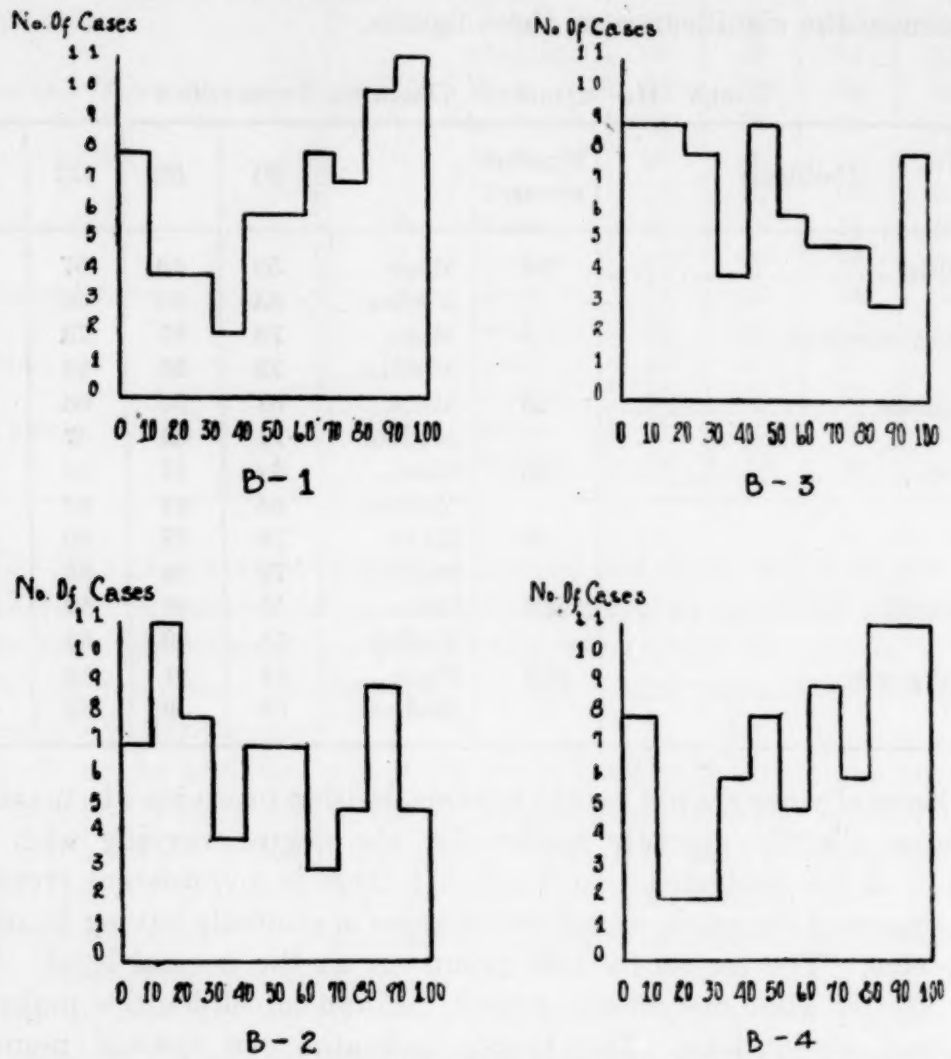

DIAGRAM 1.—Percentile scores on the four Bernreuter traits of students desiring interviews concerning social problems.

high *B*1 and *B*3 scores, indicate withdrawal, if one may assume that these diseases leave or carry with them physical effects conducive to a tendency of such nature. This assumption does not seem unreasonable. No significant departure from the norm is apparent in any other group in Table IV, possibly because the diseases incurred produced no lasting effects.

TABLE IV.—SERIOUS ILLNESS

Type of illness	Number of cases		*B*1	*B*2	*B*3	*B*4
Infantile paralysis	10	Mean	59	54	57	44
		Median	50	50	65	37
Heart trouble	31	Mean	61	54	62	43
		Median	68	62	68	39
Scarlet fever	130	Mean	51	46	51	49
		Median	51	44	51	53
Other illnesses	193	Mean	50	48	49	50
		Median	49	45	49	50
Total group	364	Mean	51	48	51	49
		Median	51	46	51	50

TABLE V.—STUDENTS HAVING CHRONIC HEADACHES AND DIGESTIVE TROUBLES

	Number of cases		*B*1	*B*2	*B*3	*B*4
Headaches	207	Mean	63	43	62	39
		Median	69	40	66	36
Digestive troubles	102	Mean	66	42	65	37
		Median	71	40	70	32
Do something about it	129	Mean	60	43	59	41
		Median	64	41	62	40
Do nothing about it	140	Mean	66	43	65	37
		Median	71	41	69	32
Total group	269	Mean	63	43	62	39
		Median	68	41	66	36

Table V seems quite significant. The relationship between neurotic tendency and headaches or digestive disturbances is well known and obvious. Note that in each instance there are indications, from the scores, of neurotic tendency, introversion, dependence, and submission. Note also that those who reported efforts to find cures or remedies are much closer to the norms than those who reported an attitude of apathy. The evidence in this table would not warrant the conclusion that the Bernreuter Inventory is a complete and accurate individual measure, but it is a strong indication that the results obtained from that Inventory are significant. Table V is consistent with what one might reasonably expect of the group under consideration.

Tables VI and VII display the scores of students who held class or certain other elective offices in high school. Throughout these scores are indicative of well-balanced personality, according to the Bernreuter norms. It is of interest—possibly of significance—to note that class secretaries recorded the lowest scores in dominance. General student government secretaries recorded higher scores.

TABLE VI.—STUDENTS HOLDING CLASS OFFICES IN HIGH SCHOOL

Office	Number of cases		*B*1	*B*2	*B*3	*B*4
President	68	Mean	45	52	44	53
		Median	45	58	40	56
Vice president	52	Mean	48	51	47	55
		Median	53	51	50	55
Secretary	47	Mean	55	39	49	47
		Median	59	27	48	49
Treasurer	38	Mean	46	48	47	55
		Median	50	46	43	53
Other offices	73	Mean	44	54	46	60
		Median	43	52	44	63
Total group	278	Mean	47	49	46	55
		Median	49	48	45	56

TABLE VII.—STUDENTS HOLDING OFFICES IN STUDENT GOVERNMENT IN HIGH SCHOOL

Office	Number of cases		*B*1	*B*2	*B*3	*B*4
President	15	Mean	43	56	48	59
		Median	45	63	45	58
Vice president	17	Mean	47	46	47	56
		Median	45	45	45	56
Secretary	12	Mean	43	48	38	61
		Median	40	40	30	63
Treasurer	8	Mean	53	55	53	49
		Median	60	55	60	45
Others	137	Mean	43	59	43	58
		Median	42	59	42	59
Total group	189	Mean	44	56	44	59
		Median	43	56	43	59

This should not be surprising to any one who knows the work usually performed by such officers. As one would expect, these students

display no marked tendency toward neuroticism or introversion, and they appear to be self-sufficient, slightly dominant.

TABLE VIII.—STUDENTS HAVING EARNED MONEY PREVIOUSLY

How much	Number of cases		*B*1	*B*2	*B*3	*B*4
Bought clothes	284	Mean	51	52	51	51
		Median	49	51	48	51
Spending money	403	Mean	49	49	49	52
		Median	49	47	48	52
Board at home	56	Mean	45	62	46	60
		Median	38	67	42	63
Board out of home	42	Mean	46	63	50	56
		Median	39	67	50	58
Occasional contributions	147	Mean	46	56	47	56
		Median	44	57	44	60
Total group	932	Mean	49	52	49	53
		Median	47	52	48	54

Table VIII shows the scores of students who earned money prior to entering the university. These scores are consistent in their trends. There is no indication of neurotic tendency, and self-sufficiency is indicated, as is a tendency in the direction of dominance. Of particular interest in this table are the scores of students who paid board. Their dominance and self-sufficiency scores are high, again an instance of the agreement of these scores with what one would expect to find.

TABLE IX.—STUDENTS HAVING SPEECH DEFECTS

	Number of cases		*B*1	*B*2	*B*3	*B*4
	50	Mean	57	45	56	41
		Median	61	43	58	44

Table IX, displaying the scores of students having speech defects, indicates clearly a tendency toward neurotic tendency and introversion. The *B*2 and *B*4 scores are consistently low, probably an indication of a retreating nature.

SUMMARY

This paper has presented and analyzed Bernreuter scores of individuals who because of a history of certain factors might reasonably be expected in some degree to deviate from the norm. If the test measures personality traits the employment of such a population should indicate the fact. If the test scores exhibited no tendency at all or disagreed sharply with what one might reasonably expect, the validity of the test as a measure of personality traits might be justly questioned.

In cases in which one would expect a neurotic tendency and introversion the mean and median scores were consistently above the 50 percentile. There is no indication in the test scores of a differentiation between these traits, but there is a definite indication of the presence of a definite personality trait. Whether it be called neurotic tendency or introversion is another question. This evidence seems to substantiate Bernreuter's contentions.

In disagreement with Flanagan's insistence that the Bernreuter measures only two personality factors, the evidence presented points to the conclusion that the *B*2 and *B*4 scales are not identical, but measure different traits. There is some indication that the *B*2 scale may point out either self-sufficiency or withdrawal, but that it should be interpreted only in the light of the scores of the other portions of the Inventory, and that it needs further refinement. Flanagan has based his sociability scale partly upon the *B*2 scale. In this interpretation he reverses the sign, and calls the trait sociability, indicating the opposite of non-sociability or independence. It is difficult to see any real distinction between this trait and "withdrawal." The evidence points very definitely to this interpretation only in certain combinations of scores. "Self-sufficiency" as a desirable confidence in one's ability is apparently measured by this scale more often than is "withdrawal." Nothing in this evidence lends support to the conclusion that only two traits are measured.

The *B*4 scales were consistent with what would be expected from the nature of the several cases. The evidence indicates that it does measure dominance.

The Bernreuter scale appears to have a degree of validity as a measure of personality traits.

THE USE OF SUBCORRELATION IN THE ANALYSIS OF NON-LINEAR OR NON-HOMOSCHEDASTIC CORRELATION CHARTS

P. S. DWYER

University of Michigan

The Pearsonian coefficient of correlation is a standard summary measure of the extent of linear trend. Its absolute value must be nearly unity if it is to be very useful in making individual predictions. A relatively small value may, however, show significant group tendencies, *i.e.*, the means of certain groups may be predicted with considerable accuracy even though r is small.

The correlation coefficients obtained in many of the researches in the fields of psychology and education are too small to be used as the bases of individual predictions, but they do establish many group tendencies. These group tendencies, when they appear, are usually demonstrated to the reader by a presentation of the correlation table itself or by a listing of the means and standard deviations of the columns (or rows) of the table.

Each of these methods is unsatisfactory. If the first alternative is chosen the data are presented in crude and undigested form and considerable study is necessary to obtain a clear picture of the relations involved. Occasionally the author indicates the positions of means (and quartiles) in the various column (or row) distributions and this helps to some extent, but still there is no summary measure which can be used as a basis of comparison with other comparable data. Also, the space available for publication hardly warrants the inclusion of data in such undigested form.

The choice of the second alternative leads to a similar criticism. The computation of means and sigmas does not provide an adequate summary measure which can be used as a basis of comparison. In addition the writers who use this method have a tendency to overemphasize the significance of the group tendencies which they find, since the method is designed to emphasize the differences in the groups but it does not provide an unbiased measure of the importance of these differences.

The Pearsonian coefficient, especially if supplemented by the standard deviation of errors, is a reasonably satisfactory measure of predictive power (1) when the regression curve is linear (the means of the columns, or rows, lie on a straight line), (2) when the distribution is homo-

schedastic (the standard deviation of the rows, or columns, are the same) and (3) when the complete range of the data is used. In case the regression curve is not linear, it is customary to compute the more or less unsatisfactory correlation ratio. This measure cannot well be compared with measures based on linear assumptions nor is it adaptable to the elaborate technique which has been built up for linear distributions.

It is proper to use the correlation coefficient for a non-homoschedastic linear form if one desires only a summary measure of predictive power. Such a measure, however, is not useful if one desires to discover the variations in predictive power within the distribution. It is here suggested that these distributions may be broken up into subgroups which are approximately linear and homoschedastic and that the correlation coefficients of these subgroups, hereafter called subcorrelations, may be used in indicating the predictive power. The standard deviation of errors—an important supplementary measure of predictive power—can be computed from the subgroup. An

TABLE I.—CORRELATION CHART FOR AN ARBITRARY HETEROSCHEDASTIC DISTRIBUTION WITH THE MEANS AND STANDARD DEVIATIONS OF ITS ROWS

Class mark		−4	−3	−2	−1	0	1	2	3	4	Mean	σ
	f	4	9	36	45	47	51	43	11	4		
4	16				1	1	2	10	1	1	+1.75	1.09
3	17				1	1	8	6	1		+1.29	.90
2	35			1	3	8	13	8	2		+ .86	1.12
1	51	1	1	4	10	10	12	10	2	1	+ .33	1.61
0	39	2	3	5	5	10	5	5	3	1	− .10	1.97
−1	48	1	2	10	10	10	9	3	2	1	− .30	1.67
−2	24		1	5	9	6	2	1			− .75	1.13
−3	10		1	3	5	1					−1.40	.80
−4	10		1	8	1						−2.00	.45

illustration is given in Table I. The correlation chart indicates approximate linearity, but a decided lack of homoschedasticity, since the three middle rows have much larger standard deviations than the remaining. The extent to which linearity and homoschedasticity are attained may be discovered by an examination of the means and standard deviations of the rows as given at the right of the table.

The correlation coefficient is found to be .50 with a corresponding standard deviation of errors of 1.46. Inspection shows the correlation in the first three rows, and that in the last three rows, is better than that in the middle rows. We compute the correlation coefficients and the standard deviation of errors for these subgroups.

A notation is needed to distinguish the various subgroups. We indicate by $r(R{:}k_1, \ldots k_s)$ the correlation coefficient obtained by using the rows whose class marks are $k_1, k_2, \ldots k_s$. Thus the correlations of the three selected subgroups are $r(R{:}4, 3, 2), r(R{:}-1, 0, 1)$, and $r(R{:}-2, -3, -4)$. The selection of columns could be indicated by $r(C{:}k_1, \ldots k_s)$. The standard deviation of errors is indicated by $S_1(R{:}k_1, \ldots k_s)$.

For the three subgroups selected the results are

$r(R{:}-4, -3, -2)$	$= .48$	$S_1(R{:}-4, -3, -2)$	$= .94$
$r(R{:}-1, 0, 1)$	$= .16$	$S_1(R{:}-1, 0, 1)$	$= 1.72$
$r(R{:}2, 3, 4)$	$= .32$	$S_1(R{:}2, 3, 4)$	$= .99$

The standard deviations of errors give good indications of the variations of predictive power, but the subcorrelations, as they stand, are not so satisfactory. None of the subcorrelations is as large as the correlation coefficient itself. This decrease is due not so much to a decreased predictive power as to a decrease in range.

In case a distribution is exactly linear, *i.e.*, the means of all the rows are exactly on the line of regression, it is possible to derive a formula showing the relation between the correlation coefficient and any subcorrelation obtained by the selection of a subgroup of rows. Let r' represent the correlation, σ_1' and σ_2' the standard deviations of the x and y variables and m' the slope of the linear regression with r, σ_1, σ_2 and m the corresponding values of any subcorrelation. The equation of the line of regression is $\bar{x} = \frac{r\sigma_1}{\sigma_2}\bar{y}$ and since $m = m', \frac{r\sigma_1}{\sigma_2} = \frac{r'\sigma_1'}{\sigma_2'}$. If $\frac{\sigma_1}{\sigma_1'} = c_1$, and $\frac{\sigma_2}{\sigma_2'} = c_2'$ we have $rc_1 = r'c_2$ and $r = r'\frac{c_2}{c_1}$. It follows at once that, in the case of a linear distribution, the subcorrelations can be readily

computed from the correlation if the standard deviations of x and y are known in each case.

The selection of a group of consecutive rows has a tendency to decrease σ_2, (and hence c_2) without an equivalent decrease of σ_1, (and hence c_1). Hence the general tendency of selection of a group of consecutive rows is for a decrease in the value of r. The standard deviation of errors is not affected by such a change in range.

It is frequently possible to form other linear homoschedastic subgroups which are not composed of consecutive rows, but in this case the reduction in scale does not necessarily trend in the direction of smaller r. In the case of the above problem the additional subcorrelations were computed.

$$r[R:-1, 0, 1] = r(R:-4, -3, -2, 2, 3, 4) = .78; \qquad S_1[R:-1, 0, 1] = .99$$

$$r(R:-4, -3, 3, 4) = .87; \qquad S_1(R:-4, -3, 3, 4) = .89$$

$$r(R:-4, 4) = .90; \qquad S_1(R:-4, 4) = .89$$

The standard deviations of errors are better indicators of the predictive power of the subgroups than are the subcorrelations, although the subcorrelations are needed in obtaining the prediction equations (the original prediction equation could be used in the case of absolute linearity) and in estimating the standard deviations of errors. The fairest measure of correlation to use as a basis of comparison with $r = .50$ would be obtained by correcting the subcorrelations for their restricted range. We assume a distribution which is the linear homoschedastic extension of the subdistribution and has its y standard deviation equal to that of the original distribution. The resulting values, which we characterize as corrected subcorrelations, are sometimes larger and sometimes smaller than the subcorrelations themselves. The assumption of homoschedasticity permits a more precise statement of the relationship between r' and any subcorrelation r of it. In addition to the relation $r = r'c_2/c_1$ deduced above we know that $S_1 = S_1'$ where $S_1 = \sigma_1\sqrt{1 - r^2}$ and $S_1' = \sigma_1'\sqrt{1 - r'^2}$ so that $c_1 = \frac{\sigma_1}{\sigma_1'} = \frac{\sqrt{1 - r'^2}}{\sqrt{1 - r^2}}$ and hence $r = r'\frac{\sqrt{1 - r^2}}{\sqrt{1 - r'^2}}c_2$. Solving for r' we get

$$r' = \frac{r}{\sqrt{c_2^2(1 - r^2) + r^2}}$$

This formula has been given previously, in somewhat different form, by Kelley[1] who refers to Pearson for a more general treatment.

If y is predicted from x and subcorrelations are formed by selecting columns, $r' = \frac{r}{\sqrt{c_2^2(1 - r^2) + r^2}}$ and $r = \frac{r'}{\sqrt{c_1'^2(1 - r'^2) + r'^2}}$ where $c_1' = \frac{1}{c_1}$. The formula $r' = \frac{r}{\sqrt{c_2^2(1 - r^2) + r^2}}$ is useful in extending the range of a given homoschedastic distribution since it demands only the evaluation of σ_2' of the extended distribution. It is possible to use it as a proper adjustment to the subcorrelations to make them comparable with the whole table. It gives

$$\begin{aligned} r'(R: -4, -3, -2) &= .80 \\ r'(R: -1, 0, 1) &= .36 \\ r'(R: 4, 3, 2) &= .64 \\ r'[R: -1, 0, 1] &= .67 \\ r'(R: -4, -3, 3, 4) &= .71 \\ r'(R: -4, 4) &= .73 \end{aligned}$$

The comparison of these values with the summary $r = .50$ gives a fair picture of the variation of the predictive power in the distribution, since these are values which would be obtained if the entire distribution were linearly homoschedastic according to the subgroup chosen. It is noticed that the correction sometimes results in an increase and sometimes in a decrease. This depends upon the value of $c_2 = \sigma_2/\sigma_2'$. When $\sigma_2 = \sigma_2'$, $c_2 = 1$, and $r' = r$. When $\sigma_2 < \sigma_2'$, $c_2 < 1$ and $r' > r$. When $\sigma_2 > \sigma_2'$, $c_2 > 1$ and $r' < r$. If the subgroup is composed of consecutive rows $\sigma_2 < \sigma_2'$ and $r' > r$. From this we draw the conclusions:

If any uncorrected correlation obtained from any group of consecutive rows is larger than the correlation coefficient, that increase in correlation is due to an increase in predictive power and not to a decrease in the range. If a more comparable measure of predictive power is needed this result should be corrected for range.

If any uncorrected subcorrelation obtained from any group of consecutive rows is smaller than the correlation coefficient, it does not follow that such a decrease is due to a decrease in predictive power. If, when corrected for range, the subcorrelation is still smaller than the correlation, then a decreased predictive power is indicated.

---

[1] Kelley: *Statistical Method*, pp. 223–230.

The significance of the subcorrelations deserves a word since the subgroups may be composed of a relatively small number of cases. The probable error can be computed for each subcorrelation and, if not significant, the subcorrelation should not be used. The reader is also referred to Kelley.[1]

It may appear at first that the use of subcorrelations may involve a great deal of additional work. The technique can be adjusted so that the amount of work is not hopeless. $\Sigma f d_x^2$, as well as $\Sigma f d_x$, has to be

TABLE II.—COMPUTATION OF THE SUBCORRELATIONS AND THE STANDARD DEVIATIONS OF ERRORS OF LINEAR HOMOSCHEDASTIC SUBGROUPS

d_y						$f_x =$	2	5	25	47	79	93	48	15	3
	f_1	$\Sigma f d_x$	$(\Sigma f d_x) d_y$	$f d_y$	$\Sigma f d_x^2$	$f d_y^2$	−4	−3	−2	−1	0	1	2	3	4
4	13	25	100	52	61	208					1	3	6	2	1
3	30	30	90	90	58	270				2	7	11	9	1	
2	48	45	90	96	89	192			1	3	8	24	10	2	
1	60	50	50	60	140	60			2	6	16	20	9	6	1
0	39	−3	0	0	63	0		1	5	7	13	10	2	1	
−1	33	−7	7	−33	87	33	1	1	4	8	10	5	2	1	
−2	40	−22	44	−80	106	160	1	2	9	9	9	6	3	1	1
−3	30	−4	33	−90	43	270		1	3	10	10	4	2		
−4	24	19	−76	−96	45	384			1	2	5	10	5	1	
							r	σ_1^2		σ_2^2	c_2^2		r'		S_1
All.....	317	126	338	−1	692	1577	.34	2.025		4.975	1.000		.34		1.33
−4, −3, −2	94	−14	1	−226	194	814	−.21	2.042		.652	.131		−.51		1.40
[−4, −3, −2]	223	140	337	265	498	763	.40	1.839		2.010	.404		.57		1.24
[−4, −3]	263	118	381	185	604	923	.45	2.094		3.015	.606		.54		1.29
−4, −3	54	8	−43	−186	88	654	−.45	1.608		.244	.049		−.91		1.13

computed for every row. Columns we label d, f, $\Sigma f d_x$, $(\Sigma f d_x) d_y$, $f d_y$, $\Sigma f d_x^2$, $f d_y^2$ and entries are inserted for every row. It is then a simple matter to total these for any rows desired and to substitute in the formula

[1] Kelley: *Statistical Method*, pp. 314–316

$$r = \frac{n\Sigma fd_x d_y - (\Sigma fd_x)(\Sigma fd_y)}{\{[n\Sigma fd_x^2 - (\Sigma fd_x)^2][n\Sigma fd_y^2 - (\Sigma fd_y)^2]\}^{\frac{1}{2}}}$$

The values of σ_1, σ_2, and S_1 are also obtainable from the entries indicated.

Table II illustrates the method and shows how a nearly homoschedastic but non-linear distribution can be broken up into linear trends.

Inspection does not indicate whether the row whose class mark is -2 should be included with the rows above or those below. The resulting subcorrelations show that it should be combined with the rows above it. The corresponding net decrease in the size of the standard deviation of errors verifies this conclusion.

Two regression equations can now be computed from the uncorrected correlations. In this manner the subcorrelation technique makes possible the use of linear regression equations in the case of non-linear correlation.

SUMMARY

A. The subcorrelation method is useful in analyzing correlation tables in which—

(1) Linear trend is present but there is wide variation in the standard deviations of the rows (or columns).

(2) The standard deviations of the rows are nearly equal but the means do not lie on the same straight line.

(3) Combination of (1) and (2).

B. All subcorrelations can be used in computing the regression equations and the standard deviations of errors. They should not be used as measures of the relative predictive power of the subgroup unless they are corrected for range, although in case a consecutive subgroup is formed from a linear group, the appearance of a subcorrelation larger than the correlation coefficient must be attributed to a decidedly superior predictive power of that subgroup.

The author was led to the idea of subcorrelation as a result of his study of the problem of predicting success in college. A standard predictive measure is high-school record. The correlation tables show that high-school record predicts better for the high group than it does for the lower groups as the charts are heteroschedastic and, to some extent, non-linear. The subcorrelation technique is suggested as a method of providing definite and comparable measures of such variations of predictive power. It will be used in later papers dealing with the prediction of college success.

THE EFFECT OF THE KNOWLEDGE OF MARKS ON THE SUBSEQUENT ACHIEVEMENT OF COLLEGE STUDENTS

PAUL J. FAY

DePauw University

In the majority of colleges and universities students are awarded open marks based on a letter system, such as *A*, *B*, *C*, *D*, and *F*, or based on a percentage system. In some institutions, however, actual marks are withheld from the students. They are told merely that their work is "satisfactory," "unsatisfactory," or "failure." This study is an attempt to compare objectively the effect of the open and closed marking systems on subsequent work in a psychology course. The effect was measured largely by acquisition of knowledge. Although there are other types of achievement more significant than acquisition of knowledge, this constitutes at least one important objective of any college course.

The experiment[1] was performed on a class in beginning psychology with a total enrollment of one hundred ninety-six. Seven members of the class were seniors; the rest were juniors or sophomores. For experimental purposes the class was divided into two groups; after discarding the names of students for whom data were incomplete the experimental group numbered eighty-nine and the control group ninety-six.

The two groups were equated on the following bases: (1) The percentile ranks on the American Council on Education Psychological Examination for high-school graduates and college freshmen were computed. The mean intelligence percentile of the experimental group was sixty-three with a standard deviation of 23.5. The mean intelligence percentile of the control group was sixty-three with a standard deviation of 26.5. The median percentiles of the respective groups were sixty-five and sixty-seven. (2) The scores on the first monthly test in psychology served as the second criterion. The mean score of the experimental group was 103.3 with a standard deviation of 14.0. The mean score of the control group was 102.9 with a standard deviation of 13.4.

The lecture and quiz method was used in teaching the course. Twice each week the two groups met together for lectures. Each of the two instructors in charge of the course delivered one-half of the lectures. The same textbook assignments were made to all students.

[1] Warren C. Middleton coöperated in conducting the experiment.

Once each week the entire class was divided into eight quiz and discussion sections. Each instructor conducted two experimental and two control sections.

At the end of each four weeks a test consisting of one hundred twenty-five or one hundred fifty objectively scored items was given to all the students. A final examination of four hundred objectively scored items was given at the end of the semester. Members of experimental groups were informed that they could secure their marks on each four-weeks test in terms of *A*, *B*, *C*, *D*, and *F*, and they were given their decile ranks on the test. For experimental purposes the following arbitrary areas were adopted; *A*, seven per cent; *B*, fifteen per cent; *C*+, twenty-one per cent; *C*, thirty-five per cent; *D*, fifteen per cent; *F*, seven per cent.[1]

No announcements concerning marks were made to the control group. To those students who inquired their marks were given in terms of *P*+ (satisfactory), *D*, and *F*. At the end of the first eight weeks the names of the *D* and *F* students of both groups were sent to the deans. Consequently, the actual experimental and control groups consisted of those students who received better than a *D* mark on the monthly test; they were so considered in computing the results.

As far as possible the artificial conditions of a laboratory experiment were avoided. None of the students were aware that an experiment was being performed. No attempt was made to insure that students in the experimental group would learn their marks other than the announcement in class that they were available. Some students of the group did not ask about them. This, of course, tended to minimize the differences between the experimental and control groups, but, on the other hand, it avoided artificiality alien to a classroom situation.[2]

To render the scores of the monthly tests and of the final examination comparable they were all converted into standard scores by means of Hull's formula, $X = K + SX_1$.[3] Consequently, the standard scores for each test and for the final examination have a possible range of zero to one hundred, a mean of fifty, and a standard deviation of fourteen. These standard scores will be used in presenting the results.

---

[1] The actual distribution of scores on each of the four monthly tests and on the final examination approached closely that of the normal probability curve.

[2] The marking system employed was that in effect at DePauw University. The students' marks at DePauw are recorded with the registrar in terms of *A*, *B*, *C*+, *C*, *D*, and *F*, but the only marks given to the students are *P*+, *D*, and *F*.

[3] Garrett, Henry E.: *Statistics in Psychology and Education*, pp. 282–283.

TABLE I.—MEAN SCORES OF EXPERIMENTAL AND CONTROL GROUPS BY LETTER GRADES

		First test		Second test		Third test		Fourth test		Final examination		Difference 1st − Final[1]	
		Exp.	Con.	Exp.	Con.	Exp.	Con.	Exp.	Con.	Exp.	Con.	Exp.	Con.
A Students on 1st test	Mean	73.00	70.50	67.50	61.90	70.30	63.75	69.70	62.50	74.15	60.62		
	σ_m	1.75	1.05	3.42	3.56	3.68	1.87	3.25	4.33	3.24	2.80		
Exp. $N = 9$	Diff.[2]	2.50	.....	5.60	.....	6.55	.....	7.20	.....	13.53	.....	+1.15	−9.88
Con. $N = 8$	σ_d	2.04	.....	4.94	.....	4.13	.....	5.41	.....	4.28	.....	3.68	2.99
B Students on 1st test	Mean	63.50	63.96	53.95	61.75	55.70	62.10	54.30	63.25	57.15	64.80		
	σ_m	.43	.43	3.17	2.23	2.18	3.99	2.74	2.53	1.79	2.24		
Exp. $N = 14$	Diff	.....	.46	.....	7.80	.....	6.40	.....	8.95	.....	7.65	−6.35	+ .84
Con. $N = 13$	σ_d	.....	.61	.....	3.88	.....	4.55	.....	3.73	.....	2.87	1.84	2.28
C+ Students on 1st test	Mean	57.08	56.50	55.95	57.05	54.05	53.50	51.35	52.50	52.69	53.50		
	σ_m	.56	.66	2.99	2.50	2.58	2.08	2.83	2.13	3.10	2.57		
Exp. $N = 13$	Diff	.58	.....	.....	1.10	.55	.....	.....	1.15	.....	.81	−4.39	−3.00
Con. $N = 20$	σ_d	.87	.....	.....	3.90	3.31	.....	.....	3.54	.....	4.03	3.15	2.65
C Students on 1st test	Mean	44.30	45.78	44.55	45.70	43.70	45.40	43.85	49.15	46.08	44.72		
	σ_m	.68	.80	1.63	1.98	1.60	1.78	1.90	1.30	1.20	1.78		
Exp. $N = 37$	Diff	.....	1.48	.....	1.15	.....	1.70	.....	5.30	1.36	.....	+1.78	−1.08
Con. $N = 36$	σ_d	.....	1.05	.....	2.56	.....	2.39	.....	2.30	2.15	.....	1.38	1.95

[1] Difference between the mean scores of each group on the first test and the final examination; a gain on the final examination is shown with a plus sign and a loss is shown with a minus sign.

[2] Difference between mean scores of experimental and control groups on each test and on the final examination: If the difference is in favor of the experimental group, it is recorded in the "Exp." column; if it is in favor of the control group, it is recorded in the "Con." column.

The mean scores of the experimental and control groups on each test and on the final examination are recorded in Table I; the standard error σ_m of each mean; the difference between the mean scores of the experimental and control groups on each test and on the final examination; the reliability of the difference in terms of the $\sigma_{diff.}$; the difference between the mean scores of each group on the first test and on the final examination; and the reliability of this difference.[1] These data are recorded for four groups of students: Those securing a mark of *A*, of *B*, of *C*+ and of *C* on the first test.

The mean scores are shown graphically in Fig. 1.

The average score of students receiving a grade of *A* on the first monthly test was less on the three subsequent tests. The control group, however, declined much more than the experimental group. On the final examination the experimental group surpassed its mean score on the initial test. There are sixty-two chances in one hundred that the increase is a true difference. The control group, on the other hand, secured a mean score 9.9 points smaller than its score on the first test. There are one hundred chances in one hundred that this is a true difference. The experimental section receiving a *C* mark on the first test declined slightly on the third and fourth tests, but gained 1.8 points on the final examination. There are ninety chances in one hundred that this is a true gain. The control group gained on the fourth test, but lost on the final examination. There are seventy-one chances in one hundred that it is a true loss.

Students in the experimental group who received marks of *B* or *C*+ on the first test achieved less in the subsequent tests and in the final examination. There are one hundred chances in one hundred that the loss of the *B* students in the final examination is a true one and ninety-two chances in one hundred that the loss of the *C*+ students is a true one. The *B* students in the control group gained slightly on the final examination. The *C*+ students in the control group lost on the final examination, but their decline was less than that of the experimental group.

In other words, if students securing an *A* on the first test knew their marks, they apparently put forth extra effort to retain their positions. The *C* students in the experimental group attempted to improve their standing: The *B* and *C*+ students, on the other hand, were apparently satisfied, did not unduly exert themselves, and consequently declined relative to the rest of the class.

[1] The small number of subjects in the *A*, *B*, and *C*+ groups makes the measures of reliability for their means and differences of little or no significance.

difference is in favor of the experimental group, it is recorded in the "Exp." column; if it is in favor of the control group, it is recorded in the "Con." column.

The correlation of scores on the first test with those on the final examination were .71 for the seventy-three students of the experimental group and .60 for the seventy-seven students in the control group. When students were told their marks, they were more likely to retain on the final examination the relative positions they occupied

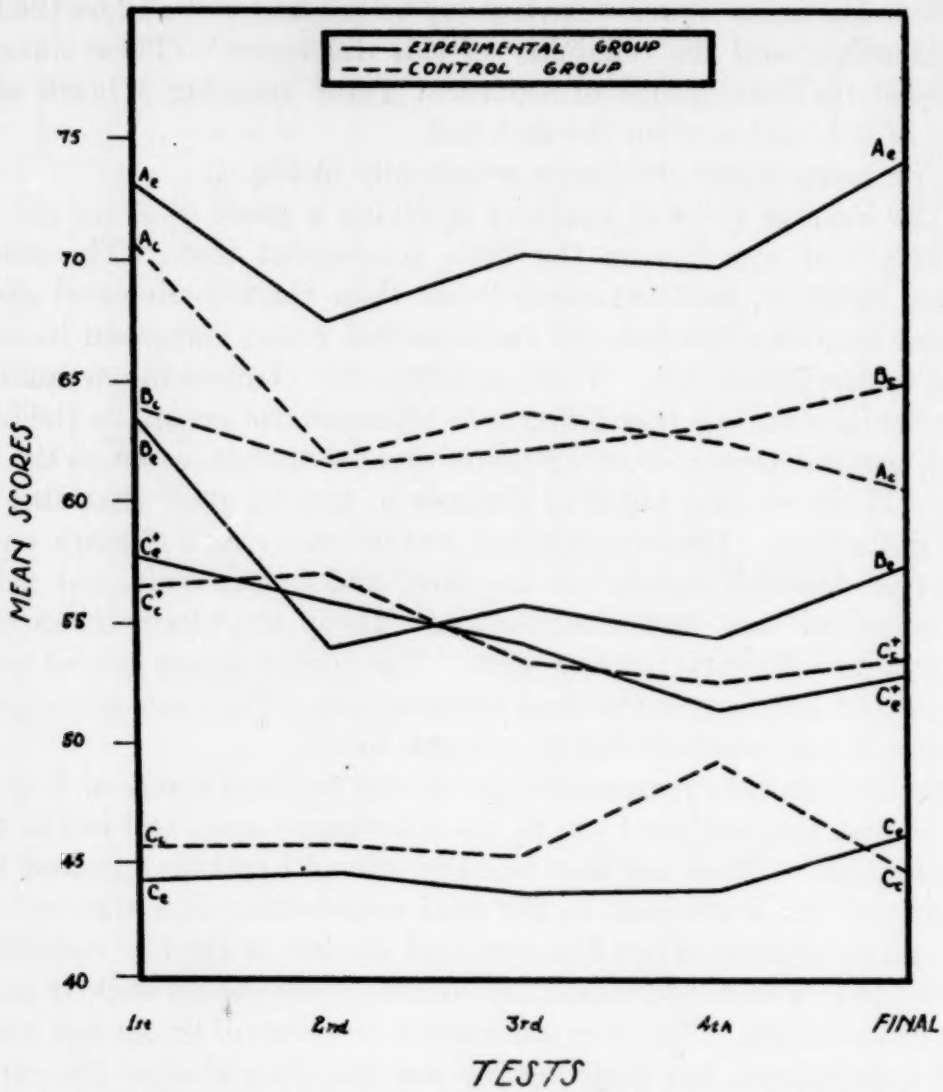

FIG. 1.—Mean scores of experimental and control groups by letter grades.

on the first test. There are eighty-nine chances in one hundred that the difference between the two correlation coefficients is a true difference. In reality, there is probably an even greater difference between them since they were both computed from partial distributions.[1]

[1] Kelly's formula, $\frac{\sigma}{\Sigma} = \frac{\sqrt{1-R}}{\sqrt{1-r}}$ is obviously inapplicable in this case (see Garrett: *Op. cit.*, p. 272). Some correction, however, would need to be made to obtain the *r*'s for more heterogeneous groups.

TABLE II.—CORRELATION BETWEEN INTELLIGENCE PERCENTILES AND GAINS ON FINAL EXAMINATIONS

Intelligence percentiles	Gain on final examination													Total
	.15–19	.20–24	.25–29	.30–34	.35–39	.40–44	.45–49	.50–54	.55–59	.60–64	.65–69	.70–74	.75–79	
	Experimental group													
90–99	..	..	..	..	..	..	3	8	1	..	..	..	..	12
80–89	..	..	..	..	1	5	1	1	1	1	1	..	..	11
70–79	..	..	..	1	..	2	4	2	1	1	..	..	..	11
60–69	..	..	..	1	1	2	4	3	1	..	..	..	..	12
50–59	1	..	..	..	..	..	2	2	2	..	1	..	..	8
40–49	..	1	..	..	..	..	1	2	3	1	..	..	..	8
30–39	..	..	..	..	..	..	2	1	1	..	..	..	..	4
20–29	..	..	..	..	..	1	1	2	1	1	..	..	..	6
10–19	..	..	..	..	..	..	1	..	..	..	..	..	..	1
Total.....	1	1	0	2	2	10	19	21	11	4	2	0	0	73
	Control group													
90–99	..	..	..	1	..	1	3	4	3	3	1	..	..	16
80–89	..	..	..	2	1	..	3	5	3	1	1	..	1	17
70–79	..	..	..	..	2	1	1	3	..	2	..	..	..	9
60–69	..	..	..	3	2	2	1	4	..	1	..	..	..	13
50–59	..	..	..	..	1	1	..	1	..	..	..	..	..	3
40–49	..	..	..	1	1	1	..	..	..	..	..	..	..	3
30–39	..	..	..	1	..	4	1	2	1	..	..	..	..	9
20–29	..	..	1	..	1	..	..	3	..	..	..	..	..	5
10–19	..	1	1	..	..	..	..	..	..	..	..	..	..	2
Total.....	..	1	2	8	8	10	9	22	7	7	2	0	1	77

Students receiving *D* or *F* on the first test were eliminated. This indicates that the *P*+ system of marking tends to result in some confusion on the part of the students as to their actual achievement.

Possibly, the most unsuspected effect of an open marking system is revealed by the correlation between intelligence percentiles and the gains on the final examination over scores on the first test. In reckoning the gains the score of each student on the first test was subtracted from his score on the final examination. Fifty was added to each of these differences to make them all positive numbers. The intelligence percentiles of each group were correlated with the gains in scores on the final examination over scores on the first test. In the experimental group an r of $-.05$ was obtained. In the control group the coefficient of correlation was .43, a significant relationship. There are one hundred chances in one hundred that the difference between the two r's is a true one.[1] The correlation tables for the two groups, shown in Table II, reveal the reason for this difference. In the control group the students of lower intelligence lacked the incentive of knowledge of their marks. Almost none of them increased their scores on the final examination. The higher intelligence of the other students helped keep them in the higher-score areas. In the experimental group, on the other hand, the students of lower intelligence increased their scores on the final examination to as great an extent as did the more intelligent students.

Further study with larger numbers of subjects is needed before any definite conclusions regarding the comparative effect of the open and closed marking systems can be confidently stated. In spite of the obvious unreliability of the results in this study, the following tentative conclusions, however, are offered:

(1) *A* students tend to achieve much more throughout the semester, if they know their marks.

(2) *B* students tend to achieve much less during the semester, if they know their marks.

(3) *C* students tend to achieve slightly more on the final examination, if they know their marks.

(4) A closed marking system tends to confuse students as to their actual achievement.

(5) Students of lower intelligence particularly need a knowledge of their marks as an incentive to increased achievement.

---

[1] This difference likewise would probably be greater if more heterogeneous groups had been used.

BOOK REVIEWS

CHARLOTTE BUEHLER and HILDEGARD HETZER. *Testing Children's Development from Birth to School Age.* Translated by Henry Beaumont. New York: Farrar and Rinehart, 1935, pp. 192.

For the past decade considerable interest has been manifest in devising scales for the measurement of behavioral development of pre-school children. Gesell is probably to be given credit for the earliest systematic attempt to make such an instrument. Shortly after the appearance of Gesell's work a series of "Baby-tests" appeared from Charlotte Buehler's laboratory at the University of Vienna. In the present volume these early tests have been thoroughly revised and extended to cover the period from one month to the sixth year of life. The present translation by Beaumont is from the first German edition of 1932.

The test items are designed to indicate behavior in six areas: Sensory reception, bodily movement, social behavior, learning, manipulation of materials, and mental production. Obviously each of these could not occur at all levels, but, for example, manipulation of materials presupposes adequate bodily movements as well as sensory reception; therefore, successful performance at the higher level indicates success at lower ones. There are seventeen groups of ten items each, one group for each of the following: Each month from one to eight months, for each two months from nine to twelve months, for each three months from thirteen to eighteen months, one for the last half (nineteen to twenty-four months) of the second year, and one group for each year from three to six. A quantitative score, the DA (developmental age) is secured by adding to, or subtracting from, the basic age (lower limit of child's CA group). Successes in the two groups above the basic age are added and failures in the two groups below the basic age are subtracted. Each item of the series has a value in terms of days. The DA may be considered analogous to the MA and may be used to calculate a DQ. While the quantitative score has a value in clinical use it is also necessary to interpret the child's performance qualitatively. The second half of the book constitutes a manual for the test. Each item is carefully described with directions for presenting it and with successful responses adequately illustrated.

In the first half of the book Buehler has presented a very penetrating analysis of the problem of testing pre-school children. Perhaps her most significant innovation is in relation to consistency of test

situation from one child to another. Most available tests are supposed to be presented in exactly the same manner to every child. Some authors have gone so far as to say that even the room in which the testing is done should be always the same. Against this practice Buehler presents very weighty arguments. Rather than an objective similarity, she feels that a functional similarity is best. Thus, she feels that testing of pre-school children should be done in their own home or, if they live in an institution, in their own room, ward, or cottage. Also, she feels that certain substitutions in materials for various test items are allowable. Wherever possible materials that the child is familiar with, his own playthings in fact, should be used. It is impossible to cite detailed examples, but her whole discussion makes this point of view so clear that it might well be studied by all psychological examiners working with very young children.

The book has many illustrations of test materials and test situations. There is also a very valuable bibliography of three hundred fifty titles including much of the work on testing pre-school children, especially that which has appeared in German. C. M. Louttit.

Indiana University.

Noel B. Cuff. *Educational Psychology.* Louisville, Ky.: Standard Printing Company, 1936, pp. VIII + 387.

Cuff's *Educational Psychology* is an eminently teachable book constructed to meet the needs of a certain group of students. It contains the standard subject-matter of educational psychology, well selected, clearly organized, and simply written. Although it possesses little striking originality, the book is by no means an uncritical summary of out-of-date theories. For example, the explanations of learning are reviewed from the laws of association of the Greeks to the recent *alpha, beta* and *gamma* hypotheses of Dunlap.

The fifteen chapters include the topics of the field of educational psychology; heredity and environment; growth and development; incentives and motives; feelings, attitudes and emotions; mental hygiene; intelligence; individual differences; three chapters on learning; transfer of training; the measurement of learning; reasoning and imagining; and socialization and guidance. Each chapter is followed by both essay and objective-type exercise material. The bibliographies have the unusual feature of designating the "preferred references" of the author's students.

This textbook should prove a useful tool of instruction in teachers colleges or wherever an elementary treatment is required.

LAURENCE F. SHAFFER.

Carnegie Institute of Technology.

JOHN EISELE DAVIS and WILLIAM RUSH DUNTON, JR. *Principles and Practice of Recreational Therapy for the Mentally Ill.* New York: A. S. Barnes and Company, 1936, pp. 197.

That physical and occupational therapy can be of aid in the treating of the mentally ill has been recognized for a long time by many workers in the field. All too often in the past, however, this type of therapy has been done in a relatively fragmentary manner. The need for coördinating such therapy with institutional programs and with individual psychotherapy has been much more often expressed than practiced. Resocializing recreational therapy in theory and practice, as described in this book, represents the most integrated approach to these problems that has come to the attention of the reviewer. In this volume a physician who has helped develop the concepts of the therapy and a physical director who has carried out the program show evidence of not only appreciating felt needs, but also of appreciating how to carry out programs that are in accord with progressive thinking in education and dynamic thinking in psychiatry.

John Dewey's *Interest and Effort in Education* is the chief source of reference for their educational thinking, and the writings of such men as Shepherd Franz, Adolph Meyer, White, are among the chief figures in psychiatry who have guided the thinking of these authors in the developing of their therapeutic schedules. The authors do not limit themselves to physical activities and the more active aspects of recreational therapy. They also consider passive recreational forms of therapy. The general purpose is expressed by them in the following sentence: "This active concept goes further than the dynamic viewpoint in that the patient is viewed not only as a modifiable social entity, but is *stimulated* into active and progressive social relationships and evaluated from the composite picture of active and progressive participation as well as the levels of inactivity which go to make up the total configuration." Recreational therapy is thus used by them as an aid in conditioning the patient to be in a state of readiness for effective psychotherapy, and in cases of senility is aimed at the postponement of deterioration.

Some of the more important subjects considered in this book are: Types and disease entities, education and reëducation, tests used, and exercises. Some of the more detailed topics included are: Play as an educative technique, possible transfer values of recreation, convalescent therapy for the mentally ill, classification of games as to complexity, classification of activities as to form, tests of mental deterioration, and the therapeutic values of work. An over-critical person might find fault with the simplified treatment of the introvert and extrovert and the specific forms of recreational therapy given for each one. Kretschmer's name is for some reason consistently misspelled. It is spelled Kritchmere. All in all, however, this is far and away the most informed and well-presented outline of the thinking and scheduling that goes to make up a comprehensive and adequate program of recreational therapy for the mentally ill. The chief value of the book is for workers in the field of physical and recreational therapy who work with the mentally ill. It has interest also for psychologists, psychiatrists, and social workers who are concerned with similar problems.

H. Meltzer.

Psychological Service Center, St. Louis.

Carter V. Good, A. S. Barr and Douglas E. Scates. *The Methodology of Educational Research.* New York: D. Appleton-Century Company, 1936, pp. XXI + 882.

About ten years ago the head of a large School of Education was approached by a member of his staff with the suggestion that a course be offered in the methodology of educational research. "But," queried the head, "what is there to put into such a course?" In a book of well over eight hundred pages the authors of *The Methodology of Educational Research* present an impressive answer.

A brief enumeration of the principal material included reveals the extent and helpfulness of the contribution which this book makes:

1. A very clear analysis of the nature of scientific thinking;
2. An unusually thorough and helpful account of bibliographic aids in locating research and other material;
3. A concise account of educational historical method, including a detailed enumeration and discussion of legal sources;
4. A lengthy account (two hundred pages) of the normative survey method, including, in addition to a summary of fact-finding procedures, a discussion of techniques involved in frequency analysis of documentary material and in the utilization of index numbers;

5. A brief but sound analysis of experimental methods and of related statistical, case study, and causal procedures;

6. Suggestive chapters on the preparation of the research report, the evaluation of educational writings and research, and the training and supervision of research workers;

7. Numerous summaries of writings on needed research in the several fields of education.

Although the book is an integrated treatise in itself, it also serves as a guide to special studies and techniques through the medium of numerous footnote and bibliographic references. If here and there the reader disagrees with a particular point made, as, for example, the importance assigned to documentary frequency studies, it must be remembered that the authors, although presenting a critical view of their own, have seemingly endeavored to mirror the various techniques actually employed in educational research. Every informed specialist in the field of education is cognizant of the discrepancy between many of the procedures commonly employed and the requirements of sound objective, valid, and significant thinking. The fact remains that regardless of the opinion that one may have concerning the quality of present-day research procedures, no one who has the slightest faith in the research approach to educational problems can afford to pass by this meaty and carefully prepared treatise.

The book contains such an amazingly large and stimulating collection of valuable material that it would seem that little remains to be done in the field of research methodology. However, with generous modesty the authors suggest that "the producer of research may need additional advanced training" in certain specialized fields. They cite the statistical analysis of data and the construction of measuring instruments as fields requiring such training. As if in response to this suggestion there has recently appeared *The Scientific Study of Educational Problems*, by W. S. Monroe and M. D. Engelhart, a book which emphasises the statistical and measurement aspects of educational research. As yet unpublished is any fully adequate advanced treatment, as far as the reviewer can judge, of the methodological issues pertaining to philosophical, evaluative, and applicational aspects of educational problems. If the authors of *The Methodology of Educational Research* have not made a perfect score in describing a complete system of research methodology, nevertheless they have set a high standard for future writers.

H. H. Abelson.

College of the City of New York.

SMILEY BLANTON and MARGARET GRAY BLANTON. *For Stutterers*. New York: D. Appelton-Century Company, 1936, pp. 186.

Stutterers can help themselves and in very mild cases they can cu themselves without professional help. In more serious cases they ca still help themselves by taking a non-resistive and coöperative attitud towards the professional. In this very intelligible little book writte by the Blantons the concepts that they use in studying and treatin stuttering are explained. The concepts used are psychoanalytic i nature. The level of presentation is simplified—perhaps, eve over-simplified.

The content of almost half of the book is concerned with genera concepts which underly the development of any emotional disorde considered from a psychoanalytic point of view. The reason for th inclusion of these concepts in this book, of course, is that they ar definitely applicable in studying and treating stuttering. Illustrativ of such ideas are: The rôle of anxiety, wishful thinking, sublimation repression, rationalization, transference of emotion. Also considered for similar reasons are growth of conscience in relation to feelings o guilt, and the significance and nature of family and social patterns. The unconscious is described as a level of mind and is treated almost naively as a storehouse.

The foregoing notions are described as being basic to an understanding of underlying causes. Stuttering itself is treated as an emotional disorder, a form of psychoneurosis. By and large the treatment of it is psychogenetic. Emphasized in their conception of stuttering are such statements as the following: That it begins with the beginning of speech that it is not inevitably outgrown; that at ten and again at the end o adolescence it significantly either diminishes or becomes worse; tha it is fundamentally a symptom and not infrequently when it disappears it shows up again in a similar blocking in thinking.

Small chapters on what parents and teachers as well as stutterers themselves can do about the problem are included. This book in no sense represents either a comprehensive or a critical study of the problem of stuttering and is not intended to do so. Its defined purpose of helping the stutterer it fulfills fairly well in spite of its over-simplification and its naive psychoanalytic assumptions. H. MELTZER.

Psychological Service Center, St. Louis.